# NAVAL POSTGRADUATE SCHOOL
## Monterey, California

DTIC
ELECTE
MAY 1 8 1995
S D
F

# THESIS

**A COMPARATIVE PRODUCT EVALUATION OF UEDIT2 VERSUS EXCEL WITH APPLICATION TO THE DHRSC SURVEY OF HEALTH CARE PREFERENCES IN THE MONTEREY CATCHMENT**

by

John H. Barnet Jr.

March 1995

Thesis Advisor: Peter A.W. Lewis

**Approved for public release; distribution is unlimited**

19950517 061

DTIC QUALITY INSPECTED 5

# REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

| 1. AGENCY USE ONLY *(Leave blank)* | 2. REPORT DATE March 95 | 3. REPORT TYPE AND DATES COVERED Master's Thesis |
|---|---|---|

4. TITLE AND SUBTITLE A COMPARATIVE PRODUCT EVALUATION OF UEDIT2 VERSUS EXCEL WITH APPLICATION TO THE DHRSC SURVEY OF HEALTH CARE PREFERENCES IN THE MONTEREY CATCHMENT

5. FUNDING NUMBERS

6. AUTHOR(S) Barnet, John H. Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

A product evaluation of the APL2-based spreadsheet editor, UEDIT2, is conducted. The program was developed at the Naval Postgraduate School by Professor Peter A.W. Lewis. A comparison of the program is made against Microsoft Excel (Version 5.0). The areas of comparison are hardware requirements, file operations, file editing, and statistics. Strengths and limitations of each program are discussed with the intent of improving UEDIT2. The main advantage of this program are its ease of use for beginning students, its ability to manipulate categorical data and perform preliminary data analysis, and its APL2 substructure. In addition, UEDIT2 provides an excellent way to prepare data for importing into prevalent statistical packages such as AGSS, SPSS, and S-PLUS.

UEDIT2 is used to conduct a data analysis on a Defense Health Resources Study Center survey comparing usage patterns of military health care beneficiaries in the Monterey region over two different time periods. Emphasis is placed on how UEDIT2 is used as an editor to perform preliminary data analysis and import data into AGSS for graphical representation. Results of the analysis indicate a reduction in military treatment facility use, but not a comparable increased reliance on TriCare sponsored health care. Little change in using private payments or private facilities has occurred between the two samples.

14. SUBJECT TERMS Text Editor, Software, UEDIT2, Spreadsheet

15. NUMBER OF PAGES 92

16. PRICE CODE

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UL |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited

# A COMPARATIVE PRODUCT EVALUATION OF UEDIT2 VERSUS EXCEL WITH APPLICATION TO THE DHRSC SURVEY OF HEALTH CARE PREFERENCES IN THE MONTEREY CATCHMENT

John H. Barnet Jr.
Lieutenant, United States Navy
B.A., University of Washington, 1987

Submitted in partial fulfillment of the
requirements for the degree of

**MASTERS OF SCIENCE IN OPERATIONS RESEARCH**

from the

**NAVAL POSTGRADUATE SCHOOL**

**March 1995**

Accesion For
NTIS CRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification .....
By .....
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1

Author: John H. Barnet Jr.

Approved by:

Peter A. W. Lewis, Advisor

William Kemple, Second Reader

Peter Purdue, Chairman
Department of Operations Research

## ABSTRACT

A product evaluation of the APL2-based spreadsheet editor, UEDIT2, is conducted. The program was developed at the Naval Postgraduate School by Professor Peter A.W. Lewis. A comparison of the program is made against Microsoft Excel (Version 5.0). The areas of comparison are hardware requirements, file operations, file editing, and statistics. Strengths and limitations of each program are discussed with the intent of improving UEDIT2. The main advantages of this program are its ease of use for beginning students, its ability to manipulate categorical data and perform preliminary data analysis, and its APL2 substructure, which allows for more advanced programming of special requirements. In addition, UEDIT2 provides an excellent way to prepare data for importing into prevalent statistical packages such as AGSS, SPSS, and S-PLUS.

UEDIT2 is used to conduct a data analysis on a Defense Health Resource Study Center survey comparing usage patterns of military health care beneficiaries in the Monterey region over two different time periods. Emphasis is placed on how UEDIT2 is used as an editor to perform preliminary analysis and import data into AGSS for graphical representation. Results of the analysis indicate a reduction in military treatment facility use but not a comparable increased reliance on TriCare sponsored health care. Little change in using private payments or private facilities has occurred between the two samples.

## THESIS DISCLAIMER

The reader is cautioned that computer programs developed in this research may not have been exercised for all cases of interest. While every effort has been made, within the time available, to ensure that the programs are free of computational and logic errors, they cannot be considered validated. Any application of these programs without additional verification is at the risk of the user.

## TABLE OF CONTENTS


I. INTRODUCTION ..... 1

A. PROBLEM STATEMENT ..... 1

B. GENERAL OUTLINE ..... 1

C. PURPOSE OF THE THESIS ..... 2

D. COMPUTATIONAL TOOLS ..... 4

II. HARDWARE AND SYSTEM COMPARISON ..... 5

A. HARDWARE REQUIREMENTS ..... 5

B. INSTALLATION AND SOFTWARE REQUIREMENTS ..... 10

III. FILE OPERATIONS ..... 13

A. OPENING AND IMPORTING A FILE ..... 13

B. EXPORTING AND SAVING A FILE ..... 17

IV. FILE EDITING ..... 21

A. DATA ENTRY ..... 21

B. DATA EDITING ..... 23

C. DATA ORGANIZATION ..... 24

V. STATISTICAL COMPARISON ..... 29

A. FREQUENCY COUNTS WITH CONDITIONS ..... 29

B. CROSSTABULATIONS WITH CONDITIONS ..... 31

C. DRAFTSMAN'S DISPLAY ..... 33

D. AGGREGATION OF CROSSTABULATIONS ..... 34

E. BOX PLOT ..... 36

F. SCATTER PLOT ..... 37

G. BOOTSTRAP ANALYSIS FOR CONTINGENCY TABLES ..... 38

H. MULTIPLE REGRESSION - ALL COMBINATIONS ..... 39

I. BOTTOM SHELF ..... 41

J. OUTPUT ..... 43

VI. RECOMMENDATIONS FOR UEDIT2 ..... 45

A. HARDWARE, INSTALLATION, AND DOCUMENTATION ..... 45
B. FILE OPERATIONS ..... 46
C. FILE EDITING ..... 49
D. STATISTICS ..... 49
E. USER INTERFACE ..... 51
VII. APPLICATION OF UEDIT2 TO THE ANALYSIS OF THE DEFENSE HEALTH RESOURCES STUDY CENTER SURVEY ON HEALTH CARE USAGE IN THE MONTEREY REGION ..... 53
A. PURPOSE OF UEDIT2 IN THE ANALYSIS ..... 53
B. BACKGROUND ON MILITARY HEALTH CARE IN MONTEREY ..... 53
C. ANALYSIS OBJECTIVE ..... 54
D. THE SURVEY ..... 55
E. ORGANIZATION OF THE DATA WITH UEDIT2 ..... 56
F. RESPONSE ..... 58
APPENDIX A. UEDIT2 AND EXCEL STATISTICAL FUNCTIONS ..... 69
APPENDIX B. PRE-CLOSING HEALTH CARE QUESTIONNAIRE ..... 71
APPENDIX C. APL2 CODE FOR SURVEY ANALYSIS ..... 73
LIST OF REFERENCES ..... 75
INITIAL DISTRIBUTION LIST ..... 77

## ACKNOWLEDGEMENT

I would like to extend my thanks for the guidance and patience by Professors Peter Lewis and William Kemple. Their ready advice and criticism made this a memorable learning experience. In addition, I would like to thank my wife, Julie, for her love, patience, and support in helping me complete my education.

## EXECUTIVE SUMMARY

When starting the Operations Research curriculum, students have little or no experience in deciding which software package should be used to enter, collect, and analyze data. As students progress through the curriculum, they are exposed to many statistical packages and to many spreadsheets. Yet, a student's natural inclination for data analysis is to turn to a program that he or she knows, that is readily accessible, and that will give him the answers he seeks as quickly as possible. Commercial spreadsheets are a common avenue chosen. Students often enter data and attempt some preliminary analysis in their spreadsheet of choice and then attempt to import the data into a more powerful statistical and graphics package for comprehensive analysis.

With a clear need for a "spreadsheet-like" editor to enter, display, print, and modify character and/or numeric data and to perform exploratory data analysis, Dr. Peter Lewis developed an APL2-based computer program, UEDIT2. The idea was to create a data analysis tool that allows the user to quickly collect, look at, and understand a data set, but lends itself to further in-depth analysis with a more powerful statistical package. But a fundamental question remains - how does UEDIT2 compare to a commercial spreadsheet? A comparative product evaluation of UEDIT2 versus Microsoft's Excel (Version 5.0) is conducted in order to answer this question.

Four areas of comparison are discussed: hardware and systems, file operations, file editing, and statistical functions. An analysis of the strengths and weaknesses of both programs is presented. A list of recommendations is made to improve the usability of UEDIT2.

As an example of an application of UEDIT2, an analysis of a health care survey is conducted. The Defense Health Resources Study Center maintains a database on how military health care beneficiaries use the different facilities and services available. The goal of the analysis given in this thesis is to graphically illustrate the distributions of usage patterns for doctor visits, hospital stays, and pharmacy use for two age groups over two different time periods. Results of the analysis indicate a strong drop in use of

military health care facilities and a slight increase in government-sponsored health care after the closure of Silas B. Hayes hospital. Little change in activity is seen for those people who rely on private insurance for health care.

# I. INTRODUCTION

## A. PROBLEM STATEMENT

When starting the Operations Research curriculum, students have little or no experience in deciding which software package should be used to collect and/or analyze data. It is not until the third or fourth quarter that a student is exposed to different types of statistical programs. As students progress to the later stages of the curriculum, they are exposed to even more statistical packages if a particular program is the preference of the instructor. Yet, a student's natural inclination is to turn to a program that he or she knows, that is readily accessible, and that will hopefully give him the answers he seeks as quickly as possible. Commercial spreadsheets are a common avenue chosen by students for this purpose; in recent years they have rapidly expanded their statistical capabilities in order to appeal to a broader base of users. Students often attempt some preliminary analysis in their spreadsheet of choice and then attempt to import the data into a more powerful statistical computing and graphics package for a comprehensive analysis.

Previously, with a clear need for a "spreadsheet-like" editor to enter, display, print, and modify character and/or numeric data and to perform exploratory data analysis, Dr. Peter Lewis developed an APL2-based computer program, UEDIT2. The initial work was done by Captain Uwe H. Steinfeld as part of his Master's Thesis [Ref. 1] and was used to analyze security data for PERSEREC. The idea was to create a data analysis tool that allows the user to quickly collect, look at, and gain an understanding of a data set, but lends itself to further in-depth analysis with a more powerful statistical package. But a fundamental question remains - how does UEDIT2 compare to a commercial spreadsheet? This thesis is an attempt to answer this fundamental question; the answer will clearly be dependent upon the needs of the individual user.

## B. GENERAL OUTLINE

This thesis is divided into three sections. The first section is a product evaluation of UEDIT2 for the purpose of preliminary statistical analysis pertaining primarily to

categorical data. The product evaluation is based on a software comparison of UEDIT2 versus Microsoft's Excel for Windows (Version 5.0). The second section of the paper is a list of recommendations to improve UEDIT2 from a human factors standpoint. One of the major advantages to having UEDIT2 developed on campus is that changes can instantly be implemented to the program. Suggestions are made in order to make UEDIT2 more "user friendly." The third section of the paper will analyze a data set from the Defense Health Resource Study Center (DHRSC). The DHRSC maintains a nation-wide database on usage of, and satisfaction with, military health care facilities versus civilian health care facilities. This analysis looks at the usage trends of the Monterey region over two different time periods - the calendar quarter preceding the closure of Silas B. Hayes hospital at Fort Ord and six months after the hospital closure. This data will be used as a basis for the software comparison and results of the analysis will be discussed.

## C. PURPOSE OF THE THESIS

The purpose of this thesis is twofold. The first purpose is to demonstrate a tool that will allow the student a fast, easy to use program for collecting and analyzing data. The second purpose is to make a contribution to the DHRSC survey by illustrating that a statistical package exists on campus that can be used for categorical data analysis, as opposed to contracting with outside agencies to do the work.

### 1. Illustration of a Superior Software Product for Personal Computers

Students often enter the curriculum with a preconceived notion of the type of software programs he or she will need. In addition, students are exposed to several packages while progressing through the curriculum. For a product evaluation, UEDIT2 was compared against Microsoft's Excel for two reasons:

1. Excel is a highly commercialized spreadsheet that students are often exposed to in previous tours of duties, or as part of a software bundle they obtain when purchasing a personal computer.

2. Students naturally prefer a software package that has an integration into word processing and graphic presentation packages. Excel is compatible with Microsoft Word and Microsoft PowerPoint (graphics presentations). Often, all of these software packages are sold as "software suites" such a Microsoft Office. Microsoft Office includes Excel, Word, and PowerPoint, as well as Access (a database package). Other software suites include Lotus' SmartSuite (1-2-3 (spreadsheet), AmiPro (word processing), and Freelance (graphics)), and Novell's Perfect Office (WordPerfect (word processing), Quattro Pro (spreadsheet), and Presentations (graphics)). Students are introduced in later quarters to statistical packages such as MINITAB, AGSS, SPSS, and S-PLUS, all of which have highly sophisticated graphical and statistical capabilities. In addition, MINITAB, SPSS, and S-PLUS have data entry tablets. However, in general, few students are familiar with these packages prior to starting the curriculum.

UEDIT2 is designed to act as a stand-alone statistical package for the beginner. Data can be received or entered in any form (numeric or character), quickly analyzed, and displayed in graphical or tabular form. Yet, as the user becomes more familiar with the program, UEDIT2 can access the APL2 computer language for programming and data transformation and manipulation.

**2. Increased Productivity on Large Databases**

While the focus of this thesis is on the use of UEDIT2 by students on home-based personal computers, it should be noted that UEDIT2 has been made available on the mainframe for handling of large files.

A perfect opportunity for using UEDIT2 is the analysis of the DHRSC survey. DHRSC was given a task - get statistics on 500,000 records of military health care beneficiaries and analyze them in order to make recommendations to the Office of the Secretary of Defense for Health Affairs (OSD(HA)) for input into the current year budget. The avenues of data collection were different (mail and phone surveys), and the turnaround time for the results was short (approximately 1 month). Because of the size of the database, outside agencies were hired for data collection and analysis. A tool such as

UEDIT2 could perhaps have been employed right on campus just as quickly, possibly saving thousands of dollars.

## D. COMPUTATIONAL TOOLS

All product testing of UEDIT2 and Excel as well as data analysis of the DHRSC survey was conducted on either a 80486-DX33 MHz or a Pentium 90 MHz personal computer. The 486 computer had 8 Mbytes of random access memory (RAM), a 80487 math co-processor, a SVGA video adapter and monitor. The Pentium computer, which has a built in math co-processor, had 32 Mbytes of RAM and a SVGA video adapter and monitor.

The UEDIT2 program requires the IBM APL2/32 interpreter for IBM compatible computers. "Its primary advantage is the ability to use all available random access memory (RAM) without the 640 KB limitation imposed by DOS." [Ref. 1] The interpreter can be built into a runtime version of UEDIT2 as an auxiliary processing file, allowing the program to act as a stand-alone package. The user need not have the APL232 program available or know how to use the APL2 programming language.

In addition, the author used MS-DOS version 6.22 as the computer's disk operating system. UEDIT2 requires that no terminate and stay resident (TSR) files be included in any start-up files. DOS versions 6.0 and later allow the user to easily program multiple configurations in the CONFIG.SYS and AUTOEXEC.BAT files. Earlier versions of DOS will require either a boot disk or editing of the start-up files to initiate the program. An OS/2 version of the program is also available.

It should be noted that although all of the graphics and tables are displayed in black and white in this thesis, the use of color plays a major role in the UEDIT2's display of results. Wherever possible, amplification of the results using color will be noted for emphasis. Even more sophisticated use of color in conjunction with the APL2 programming language is possible.

## II. HARDWARE AND SYSTEM COMPARISON

Hardware and system comparison between UEDIT2 and Microsoft Excel requires looking at two areas: 1) hardware requirements and limitations and 2) software requirements and setup procedures.

### A. HARDWARE REQUIREMENTS

Hardware requirements involve comparing the six basic computer components: machine processor, hard disk space, RAM, monitor, printer and mouse. A summary displaying each of the system requirements for the six components is shown in Table 1.

| Component | UEDIT2 | Excel |
|---|---|---|
| Machine Processor | IBM 286 or higher | IBM 286 or higher/Macintosh |
| Hard disk Space | 4.8 Mbytes | 10.5 Mbytes as a stand alone program |
| RAM | 512 Kbytes | 4 Mbytes minimum<br>8 Mbytes recommended |
| Monitor | EGA - 43 line mode<br>VGA - 50 line mode | EGA - 43 line mode<br>VGA - 50 line mode |
| Printer | Printer with HP emulation | Any printer with appropriate printer file |
| Mouse | Not Required | Required |

Table 1. Hardware Requirements for UEDIT2 and Excel.

1. RAM. Two dimensions of computer RAM are important in determining UEDIT2's database capability. The first dimension is workspace size. Unlike a typical commercial spreadsheet such as Excel where the number of rows and columns are known, UEDIT2's database handling is a function of computer RAM and time to convert information to an APL2 substructure. While this product evaluation is attempting to use a "typical" home computer with 8 Mbytes of RAM, the more RAM the computer

possesses, the larger the database size that UEDIT2 can handle. The reason for this "unknown" database size is due to the fact that UEDIT2 has an APL2 substructure. Therefore, the program's matrix dimension is further restricted to the type of data imported. For example, with 8 Mbytes of RAM, UEDIT2 can handle a strictly two digit numeric database of 10,000 rows by 20 columns. However, a database with character and numeric values having between two and six digits per cell entry is restricted to approximately 2,800 rows by 20 columns. High end personal computers and mainframe computers allow significantly larger database size.

The second dimension for database handling is time - how long the user is willing to wait for file loading as information is converted to an APL2 substructure. This aspect of database handling is a function of the speed of the computer's processor and UEDIT2's internal checking for whitespace between columns.

The more powerful the processor, the faster data can be transferred to RAM. File loading, file manipulation, and statistical functions should have some upper bound time to render a solution or to update its display. In order to test database size, two types of files were loaded into both UEDIT2 and Excel at six different file sizes. As a matter of productivity, an upper bound of ten minutes was established for the execution time on each database. The first type of file was a completely numeric file with twenty columns and an expanding number of rows. The second file type was the DHRSC Health survey that contained both character and numeric columns as well as column titles. This database contained 22 columns and an expanding number of rows. The goal of expanding the database was to see if RAM or time (as a function of processor size) is the deciding factor for each program. The results of the file loading are shown in Table 2.

In addition, these times are also very much a function of the data checking built into the program. While importing the data could be done quickly, most programs will not check beyond the first row for whitespace between columns. UEDIT2 is defaulted to check the first two rows plus a fixed percentage of the imported file. This enhanced

checking of whitespace explains the increased times for importing as the file size increases.

| File Type | File Dimensions (Row x Column) | UEDIT2 Time | Excel Time |
|---|---|---|---|
| Numeric | 10 X 20 | 0:02.5 | 0:02.0 |
| Numeric | 100 X 20 | 0:03.5 | 0:03.5 |
| Numeric | 1000 X 20 | 0:09.0 | 0:10.5 |
| Numeric | 5000 X 20 | 0:31.0 | 0:32.0 |
| Numeric | 10000 X 20 | 1:02.0 | 1:38.0 |
| Numeric | 16384 X 20 (max rows for Excel) | N/A | 3:11.5 |
| Character and Numeric | 1712 X 22 | 3:11.0 | 0:21.0 |
| Character and Numeric | 2712 X 22 | 6:57.5 | 0:25.0 |

Table 2. File Loading Times for UEDIT2 and Excel.

As the table results show, time never became the restriction for database size. RAM was the sole restriction. Yet, one factor that would increase the speed of file loading is to artificially expand the UEDIT2 executable file, namely UEDIT2.APL. In other words, if the UEDIT2 executable is made larger, the time necessary for conversion to the APL2 substructure is reduced. To confirm this idea, a large DHRSC database (2700 X 22) was loaded and then the UEDIT2 executable was resaved. The same file was reloaded and timed. The results of the file loading are shown in Table 3.

| Database Size | Original UEDIT2 Executable | Enlarged UEDIT2 Executable | Time Reduction |
|---|---|---|---|
| 1712 X 22 | 3:11.0 | 0:21.5 | -2:49.5 |
| 2712 X 22 | 6:57.5 | 3:51.0 | -3:06.5 |

Table 3. Loading Comparison for Original and Enlarged UEDIT2 Executable Files.

Excel's database size is also a function of RAM and co-processor, but with the added dimension of having a finite number of rows and columns built into the program. Excel is structured so that each file contains worksheets. Each worksheet has 16384 rows by 256 columns. Database size can be no larger than this finite size no matter how much RAM or what type of computer processor is used.

2. Monitor. UEDIT2 will operate with a monochrome or color monitor, supporting an EGA 43 line mode and a VGA 50 line mode. "However, color plays an important part in displaying the spreadsheet data, so a color monitor is preferred." [Ref. 1] UEDIT2 uses color in two significant ways. First, UEDIT2 can highlight data whether character or numeric, in one of seven colors. This added dimension allows the user to instantly program and evaluate categorical data much faster than regular black and white text. For example, if stock quotes were loaded into UEDIT2 and the user wanted to consolidate the data of all companies which reached a 52 week high, the program could be color coded to highlight only those companies matching that particular trait. Furthermore, a different color could be used to highlight those companies reaching an all-time high stock price.

The second way color is used in UEDIT2 is in the display of results of statistical calculations. UEDIT2's contingency tables and multiple regression calculations illustrate the benefits of using color to display results. In the contingency table, below the sum of the cross categorized data, UEDIT2 displays the standardized counts (residuals) for each cell in the table. However, residuals having a less than five percent chance of occurring are highlighted in yellow and residuals having a less than one percent chance are highlighted in red. The user can instantly analyze those values from the table that are statistically rare events. Excel simply calculates the sum of the cross tabulations and makes no use of standardized residuals.

A second instance of using color in statistical calculations is in the display of UEDIT2's built in "multiple regression - all combinations." The input is a p + 1 column matrix - the first p columns are the independent variables and the last column being the

dependent variable. This feature conducts regression on all data the user specifies by first taking each variable individually, then taking variables two at a time, then three at a time, etc. until all possible combinations of variables are considered. UEDIT2 then highlights in red the variables that have the strongest prediction value. The user can go directly to the variable combination that models the data best to conduct further analysis in UEDIT2's "multiple regression - specific combination." With a large number of variables, color allows the user to quickly and easily break down the data and model the system based on the strongest independent variables. While Excel offers a multiple regression option, the user must use trial and error to get the right combination of independent variables. No systematic way exists to test each subset of variables.

3. Printing. The developers of the two packages took widely different philosophies in designing how each program prints data. UEDIT2's philosophy is to default the program to print with a Hewlett Packard (HP) laser printer. Printing with any printer other than the HP laser jet requires use of an HP emulation mode. This option is usually not a problem for HP deskjet or inkjet printers. However, an HP emulation must be incorporated into other printer brands for correct printing to occur. This may prove difficult for the beginning student. So, an alternative must be developed whereby the use of any printer can be incorporated with little trouble while at the same time still keep the program structure of UEDIT2 simple. Possible alternatives are addressed in the recommendations section. UEDIT2 may only support one printer port.

Excel has built into its spreadsheet printer files for virtually every printer marketed. The user simply highlights his printer and the parallel port to which it is hooked. If the exact model of the printer is not found, loading the user's printer file into the print directory is done. Excel supports two different parallel printer ports.

4. Mouse. UEDIT2 does not support the use of a mouse for data manipulation. Since there are no "dropdown" menus or icons, highlighting and editing data is done strictly through the keyboard. Most database editing can be accomplished by learning five keystrokes (CTRL R, CTRL C, CTRL F, CTRL B, and CTRL U). The use of the on-

line help allows the user to see all the options available to him at anytime. Presently, since only ten statistical calculations are available in the statistics sub-menu in UEDIT2, and these ten functions require column or vector input, function keys are all that are required for data input. However, as more functions and transformation options are brought into the program and as the APL2 substrate is hidden in the program, a mouse will need to be incorporated in order to easily support the dimensions of arguments required for certain calculations.

Excel's use of a mouse is designed to offer the user a choice of the way in which data is manipulated. The user usually has one of four options for any action - with a mouse, with a keystroke, with a dropdown menu, or with an icon. With the number of options Excel offers and the range of users it attracts, the mouse provides the fastest way to edit data and input arguments for mathematical functions.

## B. INSTALLATION AND SOFTWARE REQUIREMENTS

**1. UEDIT2 Installation.** Installation of UEDIT2 requires that the user make five directories: an APL2 directory, a directory for the UEDIT2 executable (UWE), a directory for data storage (UED), a directory for storing font files, and a directory for storing the users manual. The APL2 directory is essential because it forms the substructure of UEDIT2. While the vast majority of individuals who use UEDIT2 have a fundamental understanding of DOS in order to set up directories, establishing what information goes into which directory can be confusing because of the lack of understanding of the APL2 substructure of the program. A strong recommendation is that a batch file be written for installation.

**2. Excel Installation.** Excel installation requires Windows version 3.1 or Windows for Workgroups version 3.11. The way Excel installs itself is by typing "setup" in the program manager of Windows and then the program searches your hard disk for space availability. The user then chooses which parts of Excel are going to be brought in. The user has a choice of several different options - such as loading the Analysis Toolpack

or the Excel solver. The users limitations in selecting these options upon program execution is a function of RAM and the user's need.

**3. UEDIT2 Startup Procedures**. Startup procedures in UEDIT2 have one software restriction - that no TSR or EMM386 (memory manager) files be running. Applications such as Windows or anti-virus software cannot be running during startup. Since most users have Windows boot-up automatically, this restriction means that changes to the AUTOEXEC.BAT and CONFIG.SYS files must be made. The changes to these files then becomes a function of which version of DOS is being used. Users having DOS version 5.0 and earlier are best off by making a system boot disk, allowing their original AUTOEXEC.BAT and CONFIG.SYS files on their hard disk to be left untouched. Users having DOS version 6.0 and later can program their AUTOEXEC.BAT and CONFIG.SYS files for multiple configurations. Thorough explanations and recommendations for installation procedures and boot files are made in Chapter VII.

**4. Excel Startup Procedures**. The only start-up restriction that Excel requires for software installation is that anti-virus programs be disabled in the user's AUTOEXEC.BAT file. The user types install in the Windows file manager and the program queues the user for the options desired.

## III. FILE OPERATIONS

In examining file operations of the two software packages, two specific areas need to be analyzed: 1) opening and importing a file and 2) saving and exporting a file.

### A. OPENING AND IMPORTING A FILE

#### 1. Opening a File in UEDIT2

Opening a file in UEDIT2 can be done in two different ways: opening an existing UEDIT2 file and opening an ASCII or comma delimited file.

When the UEDIT2 command screen is reached, the user has several options. For opening an existing UEDIT2 file, the user can simply type the file name (without the file extension) and the file is loaded in the UEDIT2 workspace. For a directory listing, a "?" is typed and a list of all of the ".UED" files are printed in alphabetical order. For opening files not contained in the default data directory, the user must first go into UEDIT2, and change where the program should look for the file by typing the APL2 command "PATH." One other option is to use UEDIT2's DOS shell command to copy or move the files of interest into the default data directory. The UEDIT2 command screen is displayed in Figure 1.

For opening non-UEDIT2 files, the operator has two choices - he can either open an ASCII file or a comma delimited CSV file from the File Operations menu. In opening ASCII files, "it is assumed that all fields in a data column ... have equal widths, padded with blanks if necessary."[Ref. 1] The user need only to type the path and file name to open the file. UEDIT2 then asks the user to name the APL2 matrix of the ASCII file. This action is necessary because UEDIT2 requires all imported data to be converted to an APL2 matrix acting as the substructure for the program. The program then will ask the user if "white space" is used to separate the columns, otherwise field lengths must be entered.

For opening a comma delimited file, the user simply selects the CSV option from the file operations menu to import the appropriate file. Options for separating the

columns (such as a tab, colon, or semi-colon) are not offered. However, one editing function that UEDIT2 offers is the ability to split columns. This splitting of columns can be done for any number of spaces within a column, so data divided by any delimiter other than a comma or white space can be imported and then split by however many characters the user chooses. The File Operations menu is shown in Figure 2.

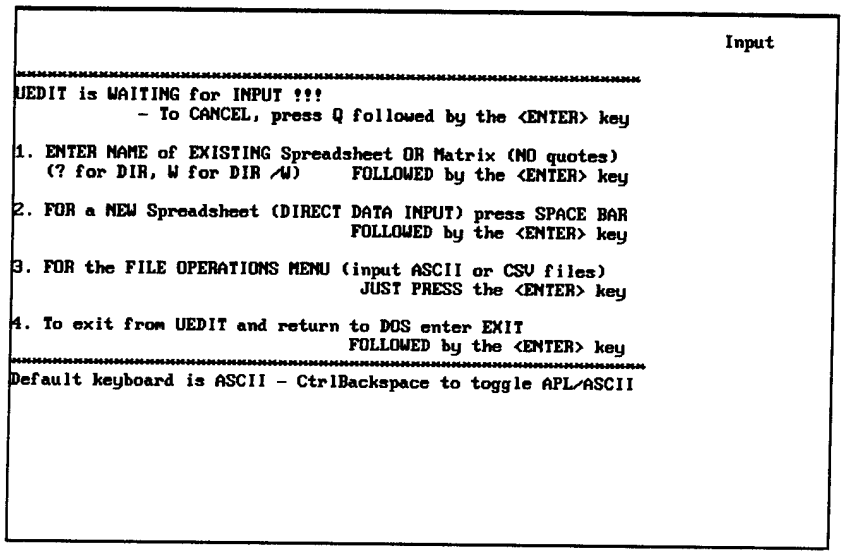



Figure 1. UEDIT2 Command Screen.

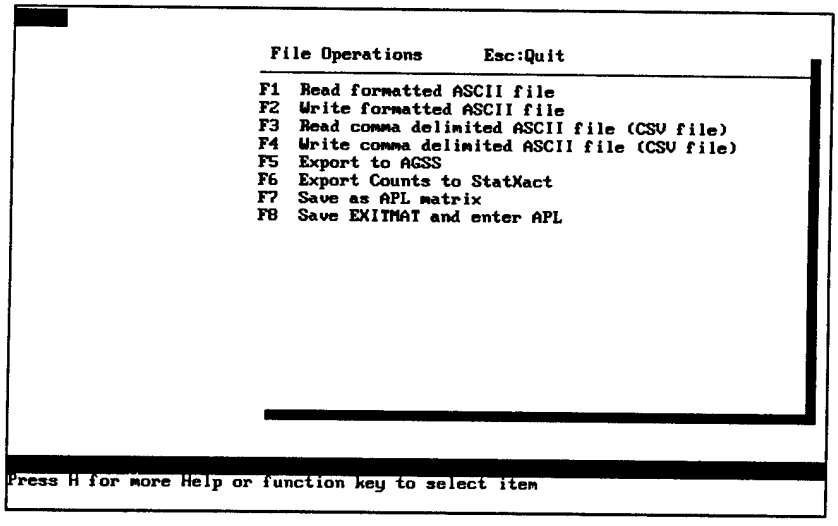



Figure 2. File Operations Menu for UEDIT2.

## 2. Opening a File in Excel

Microsoft Excel offers many different ways to open an existing file. The first way is by clicking the Icon designated for "opening." The second and third way involve the pulldown menu. The user can either click open on the file menu which will bring up the OPEN dialog box, or Excel has a feature which keeps track of the four most current worked on files. To open a recently worked on file, the user simply selects the file, regardless of its directory, and the spreadsheet will be opened. Excel also has a feature in which the user can open a file every time the program is started by storing the file in the start-up directory. The OPEN dialogue box is shown in Figure 3.

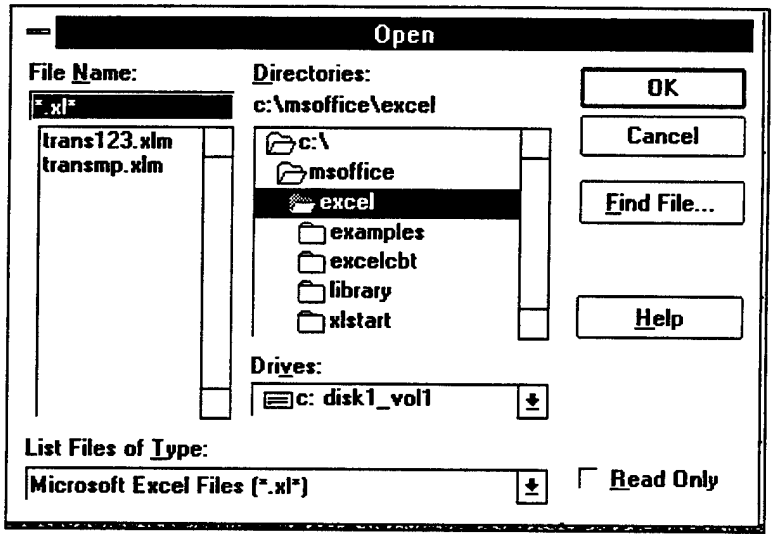



Figure 3. Excel OPEN Dialogue Box.

The OPEN file dialogue box contains several different options. The user simply selects the drive letter, directory, and file type to open. A window listing of all of the files of the particular type selected are also displayed. The user has a list of several different file types - *.* (all), text, .CSV, .PRN, as well as file extensions for Lotus and Quattro Pro. To open comma delimited files, the user has a choice of how the columns are separated - either by comma, tab, colon, or semi-colon. In addition, if a program is

being opened or created by another application, data often is imported with several columns condensed together. Having several columns of data condensed may also occur if a comma delimited file is being opened without any specification for a delimiter. Excel has an option of converting text to columns. A dialog box prompts the user with the specifications for the new columns as well as the new destination in a worksheet. Only one column of data may be converted at a time.

Several options beyond the normal OPEN dialog box are also available.

(i) The user may open several workbooks at once. This is accomplished by holding the "CTRL" key down when selecting the file names.

(ii) The user has the option of viewing the file without loading the workbook. This task is accomplished by clicking the "Read Only" option of the OPEN dialog box.

(iii) If the user cannot remember the specific name of a file, Excel has a SEARCH option. This option requires the user to make certain selection criteria of a file and then lets the program attempt to find a match. Selection criteria for search are directory or disk, specific words or phrases (a "*" may be used as a wild card), or use of summary information such as title, author name and subject from the SUMMARY dialog box that is attached to each Excel workbook. This summary information is inputted the first time the user saves a file. The SEARCH dialogue box is displayed in Figure 4.

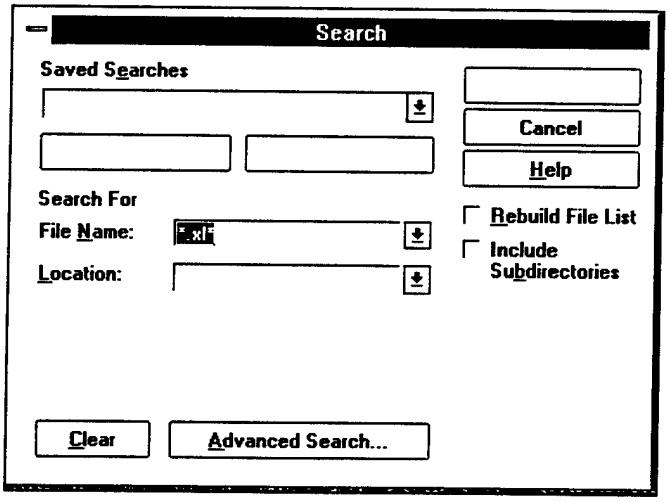



Figure 4. Excel SEARCH Dialogue Box.

## B. EXPORTING AND SAVING A FILE

### 1. Saving a File in UEDIT2

UEDIT2 offers several ways to save a file. Saving a previously unsaved spreadsheet requires the user to select the SAVE function key (F2) and enter the UEDIT2 file name. No file extension is required and the spreadsheet is automatically saved to the default data directory. All color coded, highlighted, and blocked data are retained in their original color, i.e. the .UED file is more than just the data matrix. Changing the path for saved data requires the user to implement the PATH command in the APL2 command window. In order to resave a file, the same SAVE function is employed and the user retypes the filename. In addition, UEDIT2 has a SAVE AS function which allows the user "to save a matrix under different names (with the same .UED extension) in several stages of the editing process." [Ref. 1] The newly saved matrix becomes the default name in the APL2 workspace.

In addition to saving files within the program, UEDIT2 offers several alternatives for saving data in other file formats. ASCII, CSV, AGSS and STATXACT are the formats available. Saving (writing) ASCII and CSV files are identical to opening different file types in UEDIT2. This option is started with the file operations (CTRL F2) command. It should be noted that no marked or highlighted areas are saved when exporting to a different file type.

Saving a spreadsheet in the CSV format uses a comma as the delimiter. No white space is necessary for dividing columns. "This format is supported by most commercial database programs. It is also the fastest way to import files into UEDIT2." [Ref. 1]

Exporting data to the APL2 workspace is a unique option that provides direct interface with APL2-based graphics packages such as AGSS. The users own functions can be written in the APL2 environment or in the APL2 command line of UEDIT2 with the saved vectors as arguments. If the spreadsheet is numeric data, the data is saved as a numeric vector. If the spreadsheet contains character data, a two dimensional character matrix will be formed with one character per column. Exporting data to the APL2

workspace is best illustrated through an example. For instance, say the user wants to graph the residuals of a regression problem against the model's estimate for the dependent variables. The user would highlight the vectors of residuals and model fitted variables and make an XY Scatter plot in UEDIT2. If any funneling of the data occurred, using the PUT command, data could be transformed in the APL2 environment and redisplayed in UEDIT2. Once the preliminary picture shows reasonable results, the data can be saved in APL2 and exported to AGSS for a detailed graphical display.

Saving a file in the STATXACT format specifically pertains to the crosstabulation matrix. Ordinarily, a crosstabulation will consist of just a matrix of counts. However, UEDIT2 also includes a row of residuals (standardized values) in addition to the counts. Saving a contingency table in this format "writes a matrix of counts in the crosstabulation procedure to a DOS file suitable for use in the (statistical program) STATXACT " [Ref. 1]. STATXACT is a commercial program that computes the distribution of data based on the frequency count of each cell in a crosstabulation.

Saving data in the SAVEMAT format saves the underlying APL2 matrix for use outside of UEDIT2. With numeric data, this option is synonymous with the AGSS option, except the data is saved as a matrix rather than as individual columns. The function may also be emulated in the APL2 command line by typing SAVEMAT.

While most of the save and export options are geared to IBM oriented languages, the idea is to increase the ease of programming for the users own functions. As the user becomes more familiar with the APL2 tools, these saving options significantly reduce the time to program and increase the power of UEDIT2. In addition, the ability to convert to ASCII and CSV files enable data to be incorporated to other spreadsheets, most word processors and graphics packages as text and comma delimited files.

**2. Saving a File in Excel**

Saving a document in Excel is a function of whether or not the file is a previously saved worksheet or whether the data needs to be saved in an alternate form for use in another application. Saving a workbook for the first time can be accomplished by

accessing the FILE pulldown menu or clicking the "SAVE" Icon. A dialog box is presented in which the user inputs the file name, the directory, the file type and the drive. In addition, a previously unsaved document will have a SUMMARY dialog box in which the user inputs the title of the spreadsheet, the subject, the author, keywords and summary information about the file. This dialog box serves two purposes. First, Excel uses this information as a basis for the FIND FILE option. Second, it allows the user a quick synopsis of the file. Once a file has been saved, neither the SAVE dialog box or the SUMMARY dialog box appear. Also, an autosave option can be implemented with the user selecting the time duration in which the program will automatically save a workbook with all current file settings.

Exporting data to another application requires saving the workbook in a different file format. The SAVE AS option is used and a dialog box is presented in which the user selects the file format to be saved. This dialog box is identical to OPEN dialog box for importing files.

Excel provides three special save features. First, the program has an option for creating a backup copy of each file that is saved in the same directory. The user selects how often the backup file is to be saved. Second, Excel has the option of linking and embedding files in other applications or workbooks. Linking allows "the data to be updated dynamically as the data in source document changes." [Ref. 2] This option becomes extremely useful when using the spreadsheet data in a word processor or graphics package. Excel will update the linked application if it is currently opened or the next time the application is opened. Third, Excel will allow the user to save a file upon exiting the program if editing to the workbook has occurred since the file was last saved.

# IV. FILE EDITING

Basic file functions for each program are looked at from three different areas: data entry, data editing, and data organization.

## A. DATA ENTRY

UEDIT2 offers a very easy and efficient way to enter data to a new spreadsheet or to insert data into an existing file. Using the BULK ENTRY command (F6), the program prompts the user to select row- or column-wise entry. Data is inserted above the current cursor position for row entry and to the left of the cursor position for column entry. For row-wise entry, each time a data field is entered, the cursor is automatically moved to the right. When a row is filled, the program automatically creates a new row. For column entry, the user inputs the column type and begins inputting data. When the column is completed, the cursor moves to the top of the next column, and a tone prompts the user to enter whether or not another column is needed. Additional columns may be inputted by selecting the column type and continuing. Both row- and column-wise entry are terminated at anytime by pressing the escape key.

This function has many advantages over other data entry formats. First, while many programs offer a column-wise input, few offer a row-wise input using the ENTER or RETURN key. This function allows complete entry of matrix data without the user having to look at the screen. Furthermore, UEDIT2 offers a verify function which allows the user to re-input data in a database that compares each cell entry to the original database. If a discrepancy occurs, an audible tone alerts the user to check the cell for its proper value. Second, the user may set the column widths without losing the accuracy of the data. For instance, if a number such as 123.4567 is entered in a cell, but the user wishes only for two decimal places to be displayed, 123.46 will appear in the cell of the screen, but 123.4567 will appear at the bottom of the screen in the data entry line. APL2 keeps the full accuracy of the numbers internally. In addition, as data entry occurs, the column width is automatically adjusted for the widest cell entry. The only warning the

user needs to be aware of is that data which exceeds the current column width "will appear truncated on the screen during the entering. This feature improves the speed with which UEDIT2 can handle the input and stops the input flow from being interrupted. After completion of the data entry, the necessary column widths are recalculated and the display is updated" [Ref. 1].

UEDIT2 provides two additional strengths with the bulk entry format.

(i) Missing values are automatically assigned a value. This feature is powerful because it allows the program to discard missing values for editing functions such as sorting or reducing. The program defaults missing values to -32,768 or the user may assign his own value. In addition, editing missing values allows the user to import data into graphics packages such as AGSS which otherwise would crash the program when blank values are present in the database.

(ii) No coding of the data is necessary. Text may be used in frequency counts and crosstabulations.

Excel offers many different ways to enter data. The program is defaulted to enter data in a column format upon pressing the ENTER key. Other data entry options are row-wise entry (left to right), right to left entry, bottom to top entry, and non-contiguous cell entry. Furthermore, data may be entered in more than one place in a workbook or more than one workbook at a time. All of these entry modes require pressing a unique key after each cell entry. For example, row-wise entry requires pressing the TAB key. While these functions do allow flexibility, they fail to make efficient and quick entry without the user having to check the screen.

Excel's strengths for data entry is the flexibility in its formatting. Individual cells may be formatted to unique specifications. Options for cell height and width, formatting of numbers and dates, as well as displaying formulas are available. Furthermore, individual notes may be made on each cell in order to keep track of reasons for specific values or formulas. A second strength is the program's ability to automatically fill a range of data. Ranges may be filled with the same value or with a linear or power

progression based on two data points. Date trends and custom lists may also be inserted. UEDIT2 conducts these operations, but only with APL2 commands in the APL2 command line.

In addition to different keys for different entry modes, Excel does have two additional shortfalls. First, while the user may highlight a range of data and the program will fill the range in any of the entry modes, the program does not warn the user that the range is filled. The user may inadvertently overwrite cells. Second, unlike UEDIT2, no confirmation function is available.

## B. DATA EDITING

UEDIT2 contains most of the basic editing functions found in text editors and spreadsheets. Copy and move commands are easily conducted by marking the desired area and placing the cursor at the desired spot for insertion. All actions for editing are conducted with the arrow keys since the program does not currently support a mouse. The program does have some unique features not found in other spreadsheets. For instance, UEDIT2 does support vector analysis and manipulation very easily. Functions such as the "PUT" command allow the user to place a spreadsheet vector into the APL2 substructure. Further analysis may be conducted by implementing functions from the APL2 command window in UEDIT2. Vectors contained in the APL2 workspace may also be inserted into the current file or a new file using the "GET" command.

UEDIT2 has two limitations to cutting, copying and pasting data. The first disadvantage is that the user is unable to see a current listing of all of the vectors available in the APL2 substructure. The second disadvantage is the inability to load different APL2 workspaces to work with certain files.

UEDIT2 contains three unique functions to data editing. The first is a highlighting function. While most text editors have a marking option, few posses a highlighting function. "While marking is used as a preparation for several editing functions, e.g., to copy, move, print or rotate matrix areas, highlighting is used to exploit

color to emphasize the contents of matrix fields" [Ref. 1]. Data may be highlighted in one of six different colors. A blinking "H" appears in the upper left hand corner of the screen denoting the function is in use. While highlighting will color rows, columns or blocks of data, UEDIT2 also contains a coloring function whereby different colors may be used for unique values in a column. Furthermore, this color coding is done automatically for the residuals of the crosstabulation function. Unique values are easily identified. A third use of color is the "SHADOWING" function. This function prompts the user to select a color level and then displays all highlighted data above that particular color level. The resulting matrix may be edited and saved. Unmarking, unhighlighting, or recoloring data is a one step process as is returning to the original unedited database.

Excel's strength in data editing is multiple ways to produce the same results. Each editing function may be used through a drop down menu, an Icon, a keystroke or a "short-cut" menu (the right mouse button). Marking the data with a mouse is all that is required to cut, paste, move, or copy data. Entire rows or columns may be marked by clicking the appropriate heading.

While using icons make editing simpler, Excel does have one drawback - the limited use of color. With the exception of having the option to denote negative values in red, no other use of color is made. Filtering and marking data as well as highlighting significant statistical results would aid in the speed and interpretation of data.

## C. DATA ORGANIZATION

UEDIT2 and Excel have four methods to organize and condense data. The first method is the SORT command. UEDIT2 sorts the rows of a matrix simultaneously on any number of columns in major-minor order. Character data is sorted lexigraphically and is case-insensitive. The default sorting mode is conducted in ascending order. Sorting in descending order requires a negative sign be put in front of the column number. Sorting multiple columns simultaneously is performed by first sorting the data

from the right most entered column, moving left until all columns have been completed. For example, if the user enters the columns to sort

1, 2, -3

UEDIT2 first sorts the matrix in descending order on column three, then sorts in ascending order on column two, and finally sorts on the most significant column (column one) in ascending order. Any number of columns may be sorted. Data sorted on a frequency count automatically updates the cumulative frequency. The resulting matrix is overlaid on the original data and can be edited and saved or discarded to return to the original matrix.

Excel offers similar capabilities with its sort function. Data may be sorted numerically, alphabetically, or chronologically in ascending or descending order. If duplicate data is in the first column, a second column may be selected. Sorting criteria may be made on a maximum of three columns.

Excel has three advantages over UEDIT2 for its sort function. First, data may be sorted by columns. Data is sorted in the order (ascending or descending) the user specifies. Second, partial data may be sorted. A warning is displayed before the operation is carried out to alert the user that data does not move in adjacent columns or rows. Third, since automatic subtotals can be calculated on a database, Excel offers an option to allow data to be sorted by just the subtotals. Rows or columns that are grouped together stay together.

The only disadvantage to this procedure is in returning to the original data. Excel performs its sorting on the original matrix. Therefore, if the users wishes to return to the original data after multiple sorting, before any sorting is conducted, a column must be manually inserted to label the rows in ascending order. Resorting the data according to the inserted row will return the data base to its original form. However, this method falls short if the user saves the data prior to re-sorting. If the user fails to use the Undo command at any point during the sort or saves the data, the original database is permanently changed.

A second method of organizing data is the FILTER command. Both UEDIT2 and Excel contain this feature which allows the user to quickly and easily reduce an original database to a subset of data. UEDIT2's function is called "REDUCE" and Excel's command is called "FILTER." In UEDIT2, the user simply marks the range of data and the program scrolls through each unique value. The user selects those column values with which to base the subset, and the program overlays the reduced matrix on top of the original matrix. The new matrix may be edited and saved or discarded.

In Excel, the FILTER command is performed on the original matrix. Selecting the FILTER function produces arrows next to the column labels. Clicking an arrow produces a list of the unique values in that column. The user has the option of selecting a single value, ridding the column of all blanks or displaying only non-blank entries. A custom criteria option is also available whereby the user may select two unique values or put in logical operator formulas with which to match column values. Furthermore, the user may input his own compute criteria not contained in the database but referenced in an adjacent area of the worksheet. Only rows for which an exact match is made are displayed in the submatrix. An arrow next to the column label changes from black to blue to indicate that data in the column has been reduced. Row numbers from the original matrix are kept in the reduced matrix. At the bottom of the screen, the size of the submatrix is displayed. To undo any filtering, the users selects the ALL option in each column to return the data to its original form. A reduced matrix may be copied to another location in the same file or to another worksheet altogether.

While both commands - REDUCE in UEDIT2 and FILTER in Excel - produce identical results, each has its strengths and weaknesses. UEDIT2's strength is that the original data is untampered. In addition, the user has a choice to select any number of the unique fields in the column. UEDIT2's weakness is that because of the lack of a dropdown menu, the user must scroll through every unique value in a column. Columns with many values may be cumbersome to select. It could be preferable to have a dropdown list of unique values so particular values could be selected. A second

weakness is that the user may not select a subset based on logical operators. Third, column labels are not retained in the overlaid matrix. This recommendation is being implemented in the revised version of UEDIT2.

Excel's strength is speed in selecting criteria for reducing the size of the data. Column values are displayed in ascending order for numerical values, alphabetically for text and chronologically for dates. The weakness of this function is that only a maximum of two values from each column may be selected for reduction.

Excel offers additional ways to sort and maintain lists of data. One method is called Data Form. By highlighting column labels and calling up the Data Form dialogue box, the user inputs values with which Excel compares to the database. A subset of data is created relative to the exact match or logical expression in the CRITERIA window. The user may scroll down each row of the subset and the current values for that row are displayed. The user only displays data from the columns he selects. Up to 32 columns may be displayed in a Data Form. Cell entries may be edited or protected from changes in the dialogue box. Edited cells may by cleared and returned to the original spreadsheet. In addition, new records may be added to the database directly.

A second alternative to filtering is the LOOK UP function. This feature uses a worksheet formula inputted by the user "to find one item of data that is associated with another item of data in the same list" [Ref. 2]. The user simply enters a criteria and the corresponding column values are displayed. The function may be programmed to look up an item based on the first row or first column of a database. The only restriction to this function is that input data must be in the first row or column of the database.

The third alternative to filtering is the MATCH or INDEX functions. The match function is used to find the position of a searched value when the desired column is not the first column in the database. The INDEX function is used for recalling data by an address number. The resulting cell value is displayed.

A fourth method of searching a spreadsheet is by comparing a matrix to a range of data. A range of data is labeled in ascending order. The user inputs comparative values

in ascending order and the program then compares the input to the marked range. The result will be the value or values that corresponds to the row of the input. One strength of this function is that comparative values may be returned. Having the data in ascending order allows the output to be the closest matching criteria between two rows of data. No exact match is necessary.

# V. STATISTICAL COMPARISON

Comparison of the statistical packages in UEDIT2 and Excel is best conducted by analyzing the strengths and weaknesses of UEDIT2's statistical menu and then comparing them to the strengths and weaknesses of comparable functions in Excel. Contrasts of all other statistical functions not in UEDIT2's menu but included in the APL2 substructure are also presented as part of UEDIT2's "Bottom Shelf." These functions require some basic APL2 knowledge. Statistical output and graphical displays of UEDIT2 and Excel are also discussed. A listing of all available functions for both programs can be found in Appendix A.

## A. FREQUENCY COUNTS WITH CONDITIONS

Frequency counts (histograms) are one of the most widely used statistical tools for the analysis of one dimensional categorical data. Both UEDIT2 and Excel attempt to make this function very simple for speed of analysis.

UEDIT2 prompts the user for a column on which to perform the frequency count. In addition, if the column of interest contains numeric data, the user must specify the lower bound, the upper bound, and the number of classes (bins) in which to separate the data. Frequency counts in UEDIT2 have several advantages in its implementation over Excel. First, UEDIT2 offers the feature of conditioning on any column in the spreadsheet, including the column which is being counted. This option allows the user to exclude particular values, count only data between certain dates, or count only those entries having a characteristic of another column. A second advantage is that UEDIT2 tabulates missing values. Graphical programs such as AGSS require that no missing values be present when importing data to a file or the program will crash. Having the ability to delete missing values in UEDIT2 eliminates this problem. Third, for numeric data, the number of classes to divide the data is easily created. The width of classes is equal to the upper limit of the column minus the lower limit of the column divided by the number of classes (N) specified by the user. Two additional classes are created for values

less than the lower bound and greater than the upper bound. Therefore, N + 2 total classes are created. The fourth advantage of UEDIT2 is that frequency counts are not restricted to numeric data. Categorical data may be counted and conditioned. Therefore, no recoding of text data is necessary, as it is with Excel.

Once the frequency count is displayed, along with the relative and cumulative frequencies and a simple bar chart, the table which is overlaid "on top" of the original matrix may be edited like any other UEDIT2 matrix. The user may sort the count in ascending or descending lexigraphical order. The cumulative frequencies are automatically recalculated when the resorting is done.

Only one disadvantage exists in UEDIT2's frequency count. Currently, the input into the function is designed to receive an entire column of data. Therefore, counting partial column data or a range of data is not as straightforward as counting an entire column. Partial column counting requires the user to mark the desired data and shadow all other data in the spreadsheet as intermediate steps. Counting a range of data requires the data be restructured to a single column. Allowing partial column or range counting directly is being considered for future versions of the program.

Excel's frequency count has several strengths not found in UEDIT2. The biggest advantage is that by using a mouse, the user may input full or partial row or column data as well as a range or matrix of data more easily than is done in UEDIT2. A second advantage is that the output may be in any cell alongside the data, in another worksheet, in the same file, or in another file (workbook) altogether. This feature allows for display of all the statistical data next to one another. Bin data, if not specified, is evenly divided between the upper and lower bounds of the input data. Options such as cumulative frequencies, a Pareto function which sorts the data according to bin frequency, specifying specific bin sizes, and displaying a vertical bar chart are options the user chooses when the dialogue box for histograms is presented.

Excel does have several limitations with its frequency count. First, the input range does have an upper limit of 6400 cells. Second, no conditioning on columns can be

done without filtering the data first. Third, data must be numeric. No character data may be counted. In addition, unlike UEDIT2, no method exists for re-coding text data into a numeric format. Fourth, if specific bin sizes are desired, bin sizes must be inputted by the user in some part of the worksheet and then entered as an input range in Excel's histogram dialogue box. Fifth, the output columns are not automatically formatted. The user must highlight the output and format column widths in order to read the results.

## B. CROSSTABULATIONS WITH CONDITIONS

UEDIT2 has several strong features in its crosstabulation function. First, like frequency counts, conditioning on columns is permitted. This conditioning may be done on either the columns forming the crosstabulation, or another column in the spreadsheet. Second, a Chi-squared statistic is automatically computed at the bottom of the table as well as a significance statistic which explains the dependency of the two columns being tabulated. Third, along with the marginal frequencies of rows and columns, a standardized residual is automatically calculated below each cell count in the table. The standardized residual is defined as difference between the observed cell count and the expected number for a cell assuming the two columns are independent. Numerically, if N is the total number of observations, $N_{ij}$ is the number of outcomes in the (i,j)th cell, and $N_{i\bullet}$ and $N_{\bullet j}$ are the row and column marginal values respectively, the standardized residual is

$$N_{ij} - (N_{i\bullet}*N_{\bullet j}/N) / (N_{i\bullet}*N_{\bullet j}/N)^{1/2}$$

These residuals are color coded to highlight extreme values relative to the approximation of residuals as individual standard normal random variables. Yellow indicates a value with less than a five percent chance of occurring and red indicates a value with less than a one percent chance of occurring. Extreme values are instantly located in the table. If just the frequency counts are desired for exporting to another statistical package, the user may save the table in the STATXACT mode that converts the UEDIT2 table to a table just containing the frequency counts without the residuals.

STATXACT is a commercial program that computes the distribution of data based on the frequency count of each cell in a crosstabulation. As with the frequency counts, in UEDIT2, text or numeric columns may be used.

UEDIT2's crosstabulation table has two drawbacks. The first drawback is that the table is limited to only two columns. While conditioning within a column in the table or conditioning on another column in the spreadsheet can be done, the user is unable to crosstabulate on more than two columns at a time. In addition, partial column crosstabulation requires the same data editing (marking and shadowing) as the frequency counts. The second drawback is the inability of the user to be able to sort data once the contingency table has been created. Sorting alphabetically by row or column labels or numerically by either cell values or standardized normal values would greatly enhance the display of results. This sorting of data is being considered in future versions of the program.

Excel's strength in its crosstabulation matrix (called a Pivot table) is in the formulating and editing of the table. The user highlights the data area (full or partial columns/rows) and then uses the mouse to drag the row and column labels to the desired position in the contingency table. If the user wishes to change the row or column data, this task is accomplished by using the mouse to replace the appropriate data headings. The table automatically updates. A second strength of Excel's pivot table is that, unlike the programs frequency table, text data may be counted. Third, unlike UEDIT2, more than one column of data may be included in the row or column of the table. This feature enhances the ability to display data in more than two dimensions. To add to the versatility of the table, subtotals with respect to each column or row may be displayed. Fourth, Excel allows the user to sort the data once the contingency table is formed. Sorting may be done either alphabetically by column label or numerically in ascending or descending order by values in the source data of one of the columns. A fifth strength of this table is the ability to update the contingency table. Since the pivot table output is

displayed in its own worksheet, anytime source data is changed, the user simply presses the "REFRESH" key and the program instantly updates the pivot table.

Excel's pivot table has four weakness. First, no standardized residuals are displayed in the results. Second, no color is used in the display of the results. For large tables, these limitations make it cumbersome for the user to quickly analyze the results. Third, data may only be displayed in one of two formats - either raw numbers or percentages. Fourth, while multiple columns can be inputted into both rows and columns of the table, no "on the fly" conditioning of data can occur within a column of data. Data may be edited using the FILTER command to reduce the matrix to contain only desired characteristics, but this feature would require saving this subset of data if future use is required.

## C. DRAFTSMAN'S DISPLAY

The Draftsman display is a feature unique to UEDIT2. This function is a powerful tool that allows the user to crosstabulate several columns of data at once. Contingency tables are calculated for each pair of columns. The results are laid out internally in UEDIT2 as displayed in Table 4. Each contingency table is named by the UEDIT2 program according to the column labels. Each table may be edited and saved without affecting the original data set or other created contingency tables.

| | | | | |
|---|---|---|---|---|
| $V_1$ vs. $V_2$ | $V_1$ vs. $V_3$ | $V_1$ vs. $V_4$ | ... | $V_1$ vs. $V_n$ |
| | $V_2$ vs. $V_3$ | $V_2$ vs. $V_4$ | ... | $V_2$ vs. $V_n$ |
| | | $V_3$ vs. $V_4$ | ... | $V_3$ vs. $V_n$ |
| | | | ... | ... |
| | | | | $V_{n-1}$ vs. $V_n$ |

Table 4. Draftsmen Display Matrix Format.

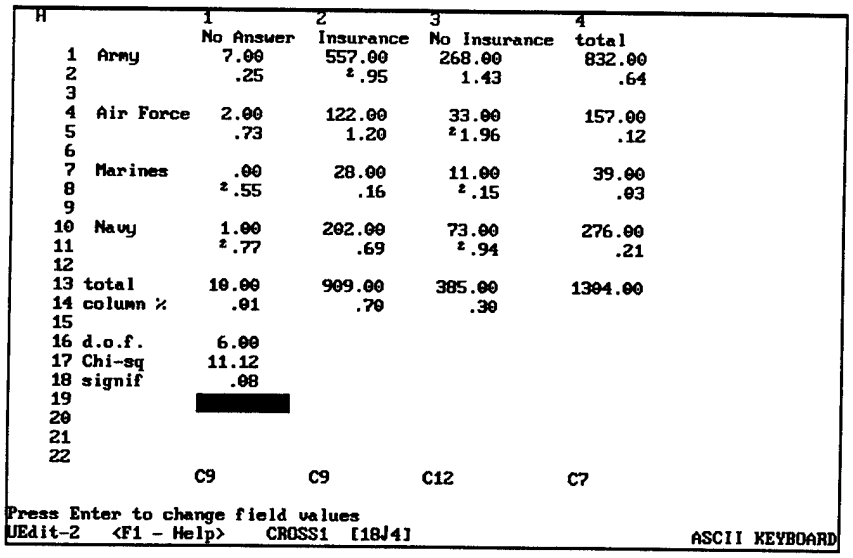
This draftsmen display has two benefits. First, the function compiles all of the column combinations automatically without the user having to input column pairs. Second, the program structures the entire set of submatrices so the user may access a specific table quickly. Numeric or character data is counted and standardized residual calculations are highlighted in specific colors to show the likelihood of occurrence. Degrees of freedom, a Chi-square statistic and significance level are automatically calculated to show the dependency of the two columns in the table.

Excel does not contain the draftsman display function. However, highlighting all of the columns a user wants to consider and using the mouse to drag specific columns into the Pivot table allows the user to obtain a contingency table for each combination. The user must keep track of the column combinations, and each new table will output into a new worksheet.

## D. AGGREGATION OF CROSSTABULATIONS

Aggregation of crosstabulations is a function in UEDIT2 that allows the user "to increase cell counts and cell expectations within the each contingency table" [Ref. 1]. Pooling of aggregated data may be done to either rows or columns and need not be contiguous. All recalculations are updated automatically, and the user always has the option of refreshing the table to its original form. For aggregation conducted in the Draftsman display mode, pooling only takes place on the submatrix being edited. All other submatrices remain unaffected.

This aggregation function is best illustrated with an example. For instance, in the DHRSC health survey, suppose one wanted to know the number of people with private insurance based on what military service an individual served. A crosstabulation could be calculated with the service column versus private insurance. The results are displayed in Table 5.

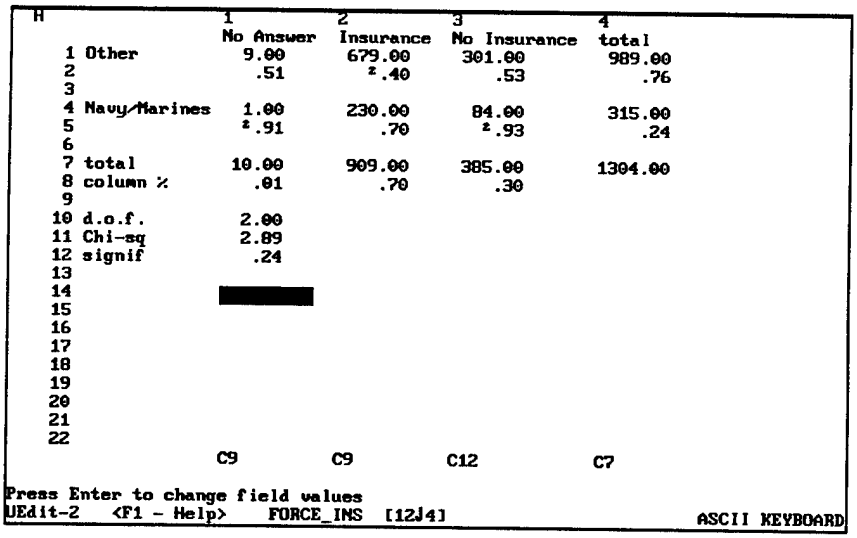


| H | | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| | | No Answer | Insurance | No Insurance | total |
| 1 | Army | 7.00 | 557.00 | 268.00 | 832.00 |
| 2 | | .25 | z.95 | 1.43 | .64 |
| 3 | | | | | |
| 4 | Air Force | 2.00 | 122.00 | 33.00 | 157.00 |
| 5 | | .73 | 1.20 | z1.96 | .12 |
| 6 | | | | | |
| 7 | Marines | .00 | 28.00 | 11.00 | 39.00 |
| 8 | | z.55 | .16 | z.15 | .03 |
| 9 | | | | | |
| 10 | Navy | 1.00 | 202.00 | 73.00 | 276.00 |
| 11 | | z.77 | .69 | z.94 | .21 |
| 12 | | | | | |
| 13 | total | 10.00 | 909.00 | 385.00 | 1304.00 |
| 14 | column % | .01 | .70 | .30 | |
| 15 | | | | | |
| 16 | d.o.f. | 6.00 | | | |
| 17 | Chi-sq | 11.12 | | | |
| 18 | signif | .08 | | | |
| 19 | | | | | |
| 20 | | | | | |
| 21 | | | | | |
| 22 | | | | | |
| | | C9 | C9 | C12 | C7 |

Press Enter to change field values
UEdit-2 <F1 - Help> CROSS1 [18J4] ASCII KEYBOARD

Table 5. Crosstabulation of Service Versus Private Insurance.

Suppose further investigation is desired on how many Navy and Marine service members had private insurance compared to the number of individuals in other services. The Aggregation of Crosstabulation function allows for table reduction from the original table. The new aggregated matrix is shown in Table 6.



| H | | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| | | No Answer | Insurance | No Insurance | total |
| 1 | Other | 9.00 | 679.00 | 301.00 | 989.00 |
| 2 | | .51 | z.40 | .53 | .76 |
| 3 | | | | | |
| 4 | Navy/Marines | 1.00 | 230.00 | 84.00 | 315.00 |
| 5 | | z.91 | .70 | z.93 | .24 |
| 6 | | | | | |
| 7 | total | 10.00 | 909.00 | 385.00 | 1304.00 |
| 8 | column % | .01 | .70 | .30 | |
| 9 | | | | | |
| 10 | d.o.f. | 2.00 | | | |
| 11 | Chi-sq | 2.89 | | | |
| 12 | signif | .24 | | | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | |
| 16 | | | | | |
| 17 | | | | | |
| 18 | | | | | |
| 19 | | | | | |
| 20 | | | | | |
| 21 | | | | | |
| 22 | | | | | |
| | | C9 | C9 | C12 | C7 |

Press Enter to change field values
UEdit-2 <F1 - Help> FORCE_INS [12J4] ASCII KEYBOARD

Table 6. Aggregated Crosstabulation Matrix.

The only drawback to this aggregation function is that if the user wishes to undo two rows or columns previously pooled, the only option available to him is to return to the original matrix. A recommendation that is being considered for future versions of UEDIT2 is to allow the user the option of being able to retract the aggregation by one step rather than having to start over on the original data set.

Excel does not offer the aggregation of columns or rows within its Pivot table. However, Excel does allow the user to filter data in the table relative to the characteristics of one or more columns. For example, if the user wishes to see a contingency table of cost versus miles per gallon on cars made in three different countries, Excel can display the table for all of the countries or for one specific country. Excel does not allow the user to display the table for two of the three countries. To conduct this option, the user must edit the source data.

## E. BOX PLOT

UEDIT2's graphics capabilities are purposely limited; the intent being to use the graphics capabilities of a companion APL2 program such as AGSS. Bar charts, box plots, and multiple variable scatter plots are the only graphics available. However, given the APL2 substructure of the program, all data may be incorporated into an APL2 graphics package like AGSS. Since UEDIT2's main purpose is to conduct fundamental categorical data analysis, the bar chart, the box plot, and the scatter plot provide adequate pictures to facilitate further numerical testing.

UEDIT2's box plot function is an easy way to display how numerical data is distributed in a column. Multiple columns may be displayed at once and the data need not be on the same scale. However, if the data is commensurate, UEDIT2 will display the box plot for each column relative to one another. The mean, median, and interquartile range are displayed within each box plot as well as symbols representing outliers (circles) and extreme outliers (stars). One recommendation under consideration is placing a tick mark at the mean to aid in identifying the shape of the distribution. A second

recommendation being considered is to display the numerical value of the sample standard deviation for each column.

One of the rewards of using the Box Plot function in UEDIT2 is an additional option called the SUBPOPULATION ANALYSIS. This function (F-10 in the Statistics menu) allows the user "to examine the frequency counts of items in a column of a spreadsheet as a function of an attribute of that population" [Ref. 1]. Attributes are selected from categorical variables in another column. The resulting spreadsheet is a series of box plots "giving the distribution of the quantity of interest for each of the subpopulations specified by the conditioning column" [Ref. 1]. A typical example of using this function would be if a column of data contained test scores and it is desired to compare the distribution of scores between male and female students.

While Excel offers a wide range of graphical options and often automatically displays a chart in the results of statistical calculations, no box plot function is available in the program. Data concerning mean, median, mode, and interquartile range may be collectively displayed using the descriptive statistics function or individually calculated by using each function option. In addition, Excel offers no subpopulation analysis option without first filtering the data.

## F. SCATTER PLOT

The object of UEDIT2's box plot function is to provide a crude visualization of the relationship between two variables to facilitate further analysis. This function's objective is not to provide a detailed picture for importing to a word processor or presentation package. Detailed graphics of the source data for presentations ought to be brought into AGSS. Therefore, the main advantage of this function is its quick and easy way of getting a simple pictorial relationship between two variables. UEDIT2 scales the dependent variable automatically and displays the results in a full screen for printing.

While not attempting to make this function a stand alone graphics package, several improvements could be made to strengthen the usability of this option. First,

UEDIT2 does not edit the missing values from the plot. Incorporating missing values can cause some very strange plots that do not represent the true relationship between variables. Currently, the user must employ the recoloring, shadowing or reduction options to the source data before an accurate picture can be portrayed. Second, the program can only display one symbol for all of the variables. Allowing for the display of more than one variable with different symbols for each variable would greatly enhance the power of this graphics function. Third, while the y-axis variable column heading is displayed at the top of the output screen, the x-axis variable has only the column number portrayed. Labeling the x-axis variable by column heading rather than column number would make the output much easier to interpret.

Since Excel is much more graphically oriented than UEDIT2, the scatter plot function is designed to be a stand alone option. Excel's scatter plot function has several strengths. First, multiple y-axis variables may be presented against a single x-axis variable. The user may also edit the type and color of symbol used to portray each variable. Trend lines, regression lines and equations may also be displayed on the graph. Second, Excel offers a 3-D option to expand the dimensions of the display. Third, the user may change the scaling, font, and labeling of each axis. Incorporating the chart into a word processor or graphics package is done simply with the cut and paste commands.

The only drawback to Excel's scatter plot function is in displaying multiple variables. Often the chart data must be edited from the source data to its own worksheet in order to get the proper presentation of the data. Little explanation is given on dividing the data into series to display proper results.

## G. BOOTSTRAP ANALYSIS FOR CONTINGENCY TABLES

This function is unique to UEDIT2 , allowing the user to simulate the distribution of data contained in a contingency table "under the assumption that the data are not associated and have marginal probabilities given by the marginal probability estimates in the table" [Ref. 1]. Since data contained in the contingency table is only a point estimate

and not necessarily a reflection of the distribution of the maximum, results can be artificially high. Therefore, by fixing the marginal totals in the table and generating cell summaries with different probabilities, the maximum value may be selected. Under the above assumption that "the data are not associated and have marginal probabilities given by the marginal probability estimates in the table," 95 or 99 percent (depending of the color coding) of the time the original calculated value should be less than the 95 or 99 percent quantile of the simulated maximums. The acceptance of the hypothesis of no dependency between columns is made if the actual value is less than the simulated quantile. This bootstrap function should be used in conjunction with the crosstabulation matrix to test the independence of columns.

Excel offers no such simulation for its pivot table.

## H. MULTIPLE REGRESSION - ALL COMBINATIONS

UEDIT2 offers a very unique way of conducting regression. Like the crosstabulation and bootstrap for crosstabulation procedures, regression is handled as a two step process. The first step is to use the MULTIPLE REGRESSION -all combinations- routine. This function allows the user to model several variables at once and then select the proper combination of variables that models the data best. All combination of variables are taken one at a time, then two at a time, etc. until every subset of variables is considered. The subsets are evaluated against various criteria for goodness of fit of the regression model, such as the sum of prediction errors (PRESS), the coefficient of determination ($R^2$), and Mallows $C_p$. The best three results for each evaluation criteria are highlighted in color (green, red, yellow respectively). The user may then take one of the specific combination of variables and input them into the MULTIPLE REGRESSION -specific combinations routine for further analysis.

This statistics function has three strong advantages over Excel. The first advantage is the ease and power of the function itself. Simply inputting several numeric columns for consideration is all that is required to model a system. UEDIT2 handles all

of the processing and presents the results in a orderly manner, complete with color indicators to quickly choose the best set of variables. Second, UEDIT2 has much stronger statistical calculations for evaluating each subset of variables. Since the $R^2$ value constantly increases as more variables are added to the model, UEDIT2 contains statistics that "penalize" the model as more variables are added. Mallows' $C_p$ and the PRESS statistic are two such examples. Analyzing these statistics allow for a tradeoff between minimizing the mean squared error and adding more variables to the model.

The second step to the regression procedure in UEDIT2 is using the MULTIPLE REGRESSION - specific combinations function. "This operation gives a detailed analysis of specific combinations of regression variables , including regression diagnostic statistics" [Ref. 1]. The function prompts for the indices of the regressor variable columns. The user has the option of including a constant term Excluding the constant term stipulates that the regression line will pass through the origin when the variables have a value of zero in the equation. The function then displays the results in spreadsheet format.

The value of this function is in the its ability to process large amounts of data quickly and organize the results for ease of interpretation. UEDIT2 offers more detailed statistics in this procedure than does Excel. Furthermore, this function may be accessed directly from the MULTIPLE REGRESSION - all combination routine. No additional input is required. The overlaid spreadsheet may be edited and sorted for the best presentation of results.

Excel has several strengths in its regression function not available in UEDIT2. First, a regression dialogue box is presented that allows the user to select the results that need to be displayed. UEDIT2 does not offer the user which results are to be displayed. A residual output table, standardized residuals, prediction values, a normal probability plot, as well as a residual plot for each variable versus its residual value may be selected. Second, output of the results may be placed next to the data, in a new worksheet, or a new file. Third, in Excel, partial ranging for input data is much easier than in UEDIT2. The

user needs only to highlight the desired data with a mouse in order to conduct calculations on a partial column. UEDIT2's partial column calculations require shadowing all unnecessary data in the spreadsheet. Fourth, a graphical display of each variable versus its residual allows the user to spot trends or flaws in the model. Multi-colinearity (high - low-high) may indicate another variable needs to be included in the model. A funneling effect may mean that the variance is dependent on the mean and the data may need to be transformed for accurate results. These errors can be much easier to identify with a graphical display.

Excel's regression function has several weaknesses. First, unlike UEDIT2, all combination of variables may not be considered at the same time. This leaves the user with having to select the variable combinations and interpret which variables are best. Selecting the combination of variables can be cumbersome as the number of variables increases. Second, the results available in the function are not as detailed as in UEDIT2. Third, the source data for independent variables must be arranged in contiguous columns for input into the regression dialogue box. For large databases, this restriction can consume large amounts of memory. Fourth, the number of input variables is limited to sixteen.

## I. BOTTOM SHELF

All further statistical functions conducted in UEDIT2 must be calculated by using a subroutine called Bottom Shelf, or by writing an appropriate APL2 function. This subroutine is part of the APL2 substructure and may be accessed through the APL2 command option (CTRL F4) in UEDIT2. A wide array of statistical functions may be employed. Univariate and multi-variate statistics, non-parametric statistics, distribution functions and significance testing are some of the basic options available. The subroutine also may conduct time series, combinatorics and reliability tests.

The major advantage of having Bottom Shelf as part of UEDIT2 is that the user is not limited to the functions provided in the statistics menu. UEDIT2 exploits the benefits

of having an APL2 substructure which is designed to conduct statistical algorithms. As the user becomes more comfortable with APL2, more powerful tools may be used for analysis

The major disadvantages of this routine are threefold. First, UEDIT2's documentation does not disclose how to use this subroutine. Without a previous knowledge of Bottom Shelf as part of the APL2 substructure, these statistical routines would not be used. Second, because of the nature of APL2 programming, beginners would have to acquire a fundamental knowledge of the language before these functions can be used. Since no basic tutorials are presented in the UEDIT2 Users Manual, accessing and inputting data for these functions are difficult. Third, results of these statistics are not presented in the UEDIT2 spreadsheet format, but rather are displayed in the APL2 workspace. To be truly beneficial to beginners, on-line help explaining each function and their results are necessary. Future versions of UEDIT2 will create a statistical menu of functions, routines, and transformations as well as on-line help to exploit the tools provided in the Bottom Shelf sub-routine. Having specific columns, ranges and parameters from a UEDIT2 database converted to an overlaid spreadsheet of results will greatly expand the power of the program.

Excel does supply a basic list of statistical calculations and functions. The strength of these functions is in the inputting of data. Highlighting the data and selecting the desired option with a mouse are all that is required. Options in the dialogue box are selected and the output may be placed next to the original data, in a new workbook, or in a new file. In addition, on-line help is easily accessible and complete with demonstrations to guide the user. Explanations are presented at each step of the examples.

The major drawback to the statistics in Excel is the lack of depth to which these functions are taken. For example, Excel can calculate the $R^2$ value between two variables but is unable to calculate the weighted $R^2$ value. In addition, time series functions, non-

parametric statistics and combinatorics are not available. The program is designed to cover the basic functions of most users but not to the detail of advanced users.

## J. OUTPUT

A key point needs to be made about the output of the statistical results. UEDIT2's statistical output is a matrix that is overlaid on top of the original matrix of data. Currently, no direct method exists to allow output to be placed next to the data or to have more than one statistical output on the same spreadsheet of results. However, the design for combining multiple outputs and aggregating files is under consideration and will be made in future versions of UEDIT2. Excel does allow the user the option of where he would like the data - either next to the original data, at a user specified address or on a new worksheet.

## VI. RECOMMENDATIONS FOR UEDIT2

In conducting the software comparison of UEDIT2 to Excel, and in using UEDIT2 in the data analysis of the DHRSC survey, several observations and recommendations are made that the author believes will improve the interface and usability of the UEDIT2 program. The recommendations are made as pertinent to the different categories of the software comparison.

### A. HARDWARE, INSTALLATION, AND DOCUMENTATION

1. As part of the ReadMe.1st file, a generic AUTOEXEC.BAT and CONFIG.SYS file should be included along with an explanation as to why most TSR functions need to be removed. Since the program is intended for home PC use and most students operate under a Microsoft Windows or DOS environment, explanation for the removal of such programs as Windows and Virus checkers needs to be mentioned. Examples for DOS 5.0 and DOS 6.0 have different ways of allowing UEDIT2 bootup. Both versions need to be explained.

2. The opening screen should be a command screen whereby the user presses a number to enact different options. The following menu is recommended:

| | |
|---|---|
| 1. | Edit UEDIT2 matrix |
| 2. | New UEDIT2 matrix |
| 3. | View .UED Library |
| 4. | Import ASCII or CSV file |
| 5. | Browse a File |
| 6. | General Help |
| 7. | Exit and Return to DOS |
| | |
| Make Selection ___ | |

3. A comprehensive index needs to be added to the user manual. In addition, an updated tutorial must be written which either incorporates all of the statistics functions through one example data set or provides small examples for each individual function.

An addition of an APL2 introduction should also be made that is separate from the UEDIT2 tutorial in which the basic APL2 commands are explained. In addition, an example of the interaction between UEDIT2 and APL2 should be presented.

4. The prompt for selecting the on-line help should be blinking. When the user selects <F1> to see the command key index, in order to receive on-line help, the user must press "H" followed by the key with which the user wants help. A blinking prompt would attract the user's attention in order to receive help.

## B. FILE OPERATIONS

1. When entering data in the bulk mode, a phrase on the screen should display "Press ESC to end Bulk Mode." In addition, the display for choosing which way to enter the data should read "Enter data down a column or across a row." Explanation in the on-line help should state that bulk mode entry at a particular cell address may be inputted by the user with the cursor position being the default start point.

2. A statement should be displayed on the screen to inform the user in the APL2 command mode, "Press F4 or ESC to end APL2 command mode."

3. When saving a file, the path and file name should be the default directory rather than having the user retype the file name every time the SAVE option is selected. For example, if an individual is working on a categorical file labeled "Test," every time the user wishes to save the file, he must re-type the name "Test" upon choosing the save option. An improvement would be to have the name "Test" appear as the file to be saved upon choosing the SAVE command. This option becomes particularly useful as the user edits data after making statistical calculations on a new spreadsheet placed over the original data. The user would no longer need to remember the name of the matrix he had been working on. In addition, allowing the user to type the path where he would like a file to be saved would allow easy saving onto a floppy disk or another directory rather than always saving to the default data directory.

4. Incorporate a summary command for each matrix and allow the user to preview these summaries before loading a file. . Since UEDIT2 must convert all data to an APL2 matrix, having a browse option would prevent the user form loading the wrong database inadvertently or force the user to remember the name of a UEDIT2 title. As UEDIT2 is a DOS based program, this addition would best be implemented as a "hidden" character column whereby the user would type a description of the saved database. This column would only be displayed upon selecting a special browse function key from the main UEDIT2 menu.

5. The ability to append new spreadsheets would be a strong addition to the program. Since UEDIT2 is a DOS based program, only one spreadsheet may be opened at a time. Having the ability to attach files to one another or being able to put together all of the submatrices created by performing statistical calculations would allow for better presentation of results.

6. Before exiting the original UEDIT2 matrix, the program should have a prompt to confirm whether or not the user would like to exit to the main menu. In addition, since file editing and statistical calculations overlap on the original data set, a display on the bottom of each screen telling which level of UEDIT2 is currently being displayed would help the user determine which matrix is being manipulated..

7. In the print command menu, when editing options such as font size or markings, after changing the different options, the program should not exit the menu. The only time the program should exit is when printing the worksheet or printing a marked section is selected. In addition, the option for landscape or portrait mode should show the current mode for printing, rather than the option to switch to.

8. When going from the UEDIT2 program to the DOS shell, the user needs a prompt as to how to return to UEDIT2. No such prompt currently exists.

9. When writing an ASCII file or a CSV file and the path the user directs the file to be written to has no disk drive, a warning needs to be installed to insert a disk and press enter to continue.

10. Allow the user to have multiple conditions be implemented when aggregating or searching a database. Currently, conducting subpopulation analysis has only one condition that the user is able to reduce the size of the data. However, if the user were allowed to take the original matrix and reduce the number of rows based on conditions of one or more columns, conducting analysis on subpopulations would be much quicker and make UEDIT2 much more versatile.

11. A tone needs to be implemented in the Save/Verify option. Since UEDIT2 allows the users to verify data in a separate matrix, and the most efficient way of re-entering the data is through the bulk entry mode, having a tone sound when the program encounters a discrepancy between the original matrix and the verifying matrix would alert the user to look at the screen to correct the cell entry.

12. When printing a document and the user inputs headings, have those headings stored in a variable so that if a reprinting of the same matrix is performed; retyping the headings are not necessary.

13. Since UEDIT2 is a DOS based text editor with an APL2 substructure, the symbols used in the graphics portion should be compatible with ASCII characters. Currently, frequency counts display a box character that, when the file is translated into a text file, are changed to different symbols. Having the graph in the UEDIT2 format the same as the text format becomes especially useful when attempting to import a file into a word processor.

14. Incorporate a mouse and "drop-down" menu structure in addition to the keyboard commands. This menu hierarchical structure allows the user to see all of the possible actions with in a particular category. A mouse makes implementing these options faster. Operations such as marking entire columns or rows could simply be performed by "clicking" the appropriate heading. While working within the program structure becomes easier as the user gains more experience with the program, both a drop-down menu and a mouse allow a much faster learning curve for beginners who

know what they want to do but do not know how to implement the operations on the keyboard.

## C. FILE EDITING

1. The program should be able to sort by columns. Currently, only row sorting is available.

2. The REDUCE function needs to save the column labels from the original matrix. Currently, no labels are displayed in the edited matrix.

3. With the RECOLOR command, have the user input which column to recolor. Presently the cursor position determines which column is colored.

## D. STATISTICS

1. In writing the on-line help for the statistical portion of the program, fuller explanation of each method and formulas need to be expanded in the manual. This requirement becomes evident in the multiple regression and specific combination regression where the formulas are written, yet no amplifying information is given. Having explanation of the relevance of the statistics and formulas incorporated in the program will enable the student to learn the background of each calculation.

2. In addition to an expanded explanation of the statistics, an expanded tutorial is needed. While the tutorial included is a good introduction, more of the statistics in the program need to be incorporated. At the same time, creating a tutorial for an introduction to the APL2 programming language would add a dimension to the program that would allow the user to expand the power of UEDIT2 if he desires. Removing the APL2 from the current tutorial and shifting it to its own tutorial with basic functions and keyboard layout is one option.

3. When selecting the conditioning option in frequency counts or cross tabulations, the program currently displays the extreme values of the conditioning column. A helpful addition to the conditioning function would be to allow the user to

condition on logical expressions of "less than" or "greater than" rather than forcing the user to input specific values.

4. Display the conditions on the Frequency Count and Crosstabulation matrix. Currently, the user must remember any conditioning he implements. Yet, if the program could display at the bottom of the new matrix the actual conditions the user selected, later referrals to the matrices would leave no doubt as to what these calculations represent. In addition, if no conditions are made, the default for the display would be "None."

5. In the Box Plot function, putting the column headings from the original matrix would enhance the newly formed matrix without having to refer to the original matrix for what the box plot is displaying. In addition, displaying the sample standard deviation next to the mean would be nice addition to the display.

6. The Aggregation statistics function should be changed to Aggregation of Crosstabulation since this is its sole function in the program.

7. When conducting multiple row or column aggregations in a crosstabulation, the user needs to have the ability to "back up" the aggregation step by step. Currently, quitting any multiple aggregation returns the user to the original matrix.

8. Incorporate a "Describe" function as an option, where such statistics as mean, median, mode, sample standard deviation, kurtosis and quartile ranges would be displayed. Adding these functions to the Box Plot display is an alternative. The Describe function should also have the ability to describe a row as well as column.

9. Have a crosstabulation option without the normalization numbers.

10. When conditioning with frequency counts or crosstabulations, an alert to the user needs to be incorporated so the user knows the program is working. Currently, the program appears as if it is locked when in fact it is working.

11. To implement statistical functions in UEDIT2 requires the user to program a vector or field value in the APL2 workspace. An alternative to this method is a function/transformation menu requiring the user to select the desired option by clicking a mouse or moving an arrow key to the desired option.

12. In the results of the ANOVA statistics, have the program display the critical values for the "F" distribution so the user can see the comparison necessary for deciding the results of the null hypothesis.

## E. USER INTERFACE

Remove the section discussing the renaming of an ALP2 matrix when importing an ASCII or CSV file. The APL2 matrix, while essential to the internal workings of the program, should have the same name as the file being imported. Explaining the renaming for the APL2 structure is unnecessary.

All recommendations suggested have either been implemented into the program or are currently under design for the latest version of UEDIT2.

# VII. APPLICATION OF UEDIT2 TO THE ANALYSIS OF THE DEFENSE HEALTH RESOURCES STUDY CENTER SURVEY ON HEALTH CARE USAGE IN THE MONTEREY REGION

## A. PURPOSE OF UEDIT2 IN THE ANALYSIS

UEDIT2 has three functions. First, the program displays, organizes and edits data. Second, the program performs preliminary data analysis. Third, UEDIT2 exports data to detailed statistical packages for more extensive data analysis and graphical display. This chapter shows how UEDIT2 is implemented in each of the three ways. The data analyzed is a survey of how military health care beneficiaries use different medical treatment facilities and services available in the Monterey area.

## B. BACKGROUND ON MILITARY HEALTH CARE IN MONTEREY

Nationally and locally, skyrocketing health care costs have become the focal point of recent debate. Nationally, health care reform has been at the forefront of the current administration's agenda, eclipsing every issue from the national deficit to welfare reform and illegal immigration. Locally, with the closure of Silas B. Hayes hospital at Fort Ord and the change in government insurance from CHAMPUS to TriCare, the health care issue has become more personalized to military families and retirees as they search for ways to solve their health care needs. It is with this prevailing issue of how best to provide health care to families in a time of shrinking budgets that ways to efficiently allocate limited funds must be employed.

Acquiring data on how military health care beneficiaries use the different treatment avenues available to them is a first step in more efficiently allocating funds. Having this data aids in solving two problems. First, the data would help decision makers allocate money more effectively to health care agencies, the theory being that those agencies receiving the most use by beneficiaries would receive the most money budgeted for a particular geographic area. Second, as a result of the first problem, the data would

help agencies who are competing for health care dollars to provide better, more efficient service to their clients. Failure to do so would eliminate the agencies services.

On a national level, current allocations for military health care are estimated at 17 billion dollars. [Ref. 3] With such a large dollar amount at stake, the Department of Defense has recognized the need to obtain a database of how its personnel use military treatment facilities (MTFs). The Office of the Secretary of Defense (Health Affairs) (OSD(HA)) has decided that as input into future defense budgets using a fixed resource amount per beneficiary to each MTF or TriCare administrator is necessary. OSD(HA)'s intent is to "identify and implement more efficient management of limited resources by eliminating financial incentives for increasing the number of services and/or providing more costly care. In other words, each medical treatment facility will assume responsibility for providing health services to a defined population, for a fixed amount per beneficiary." [Ref. 3]

In order to calculate the amount needed per beneficiary, the number of MTF users must be counted and the pattern of use by beneficiaries in a particular region must be identified. So, the OSD(HA) chartered the Defense Health Resources Study Center (DHRSC) to collect a database of military health care beneficiaries. The DHRSC began surveying opinions and preferences of different options for health care. The DHRSC forwards survey results to OSD(HA) who, in turn, makes recommendations on defense health care funding to either MTF's or TriCare in future presidential budgets. The database is updated quarterly by giving the survey to another sample of military health care beneficiaries in the same region.

## C. ANALYSIS OBJECTIVE

What the analysis in this thesis does, is look at and analyze the survey responses of the Monterey catchment area over two different time periods. A catchment area is a 40 mile radius around a MTF. The MTF in the Monterey catchment is Silas B. Hayes at Fort Ord, California. What makes these two samples so unique is that between surveys, Silas

B. Hayes hospital was closed. All beneficiaries without private insurance or opting to use MTF facilities had to seek medical care at another MTF, presumably Oakland Naval Hospital.

Each beneficiary has three avenues available for meeting his health care needs - MTFs, TriCare sponsored health care, or health care through private payment/insurance. Each of the health care alternatives are not mutually exclusive. In other words, depending on such factors as age and waiver eligibility, beneficiaries are not limited to only one avenue for health care. The emphasis of this analysis focuses on two age groups - retirees and their dependents over age 65 and retirees and their dependents under age 65. Since active duty members and their dependents were not questioned in the second sample, these categories of beneficiaries are not analyzed. The two sample responses allow for observation of the shift in military health care options as a result of the loss of a MTF. The primary purpose of the analysis is to illustrate the usage patterns of inpatient hospital care, doctor visits and pharmacy services for the two age groups and then document the percent change in use between the two time periods.

## D. THE SURVEY

An eight question survey was issued by mail. Respondents had the choice of answering by mail or by phone. Surveys answered by phone were stored in a database file at the Naval Postgraduate School (NPS) that was updated with each response. DHRSC downloaded the data from the Mainframe computer at NPS for analysis. Surveys answered by mail were sent to the Defense Eligibility and Enrollment Reporting System (DEERS) Support Office , where they were scanned into a database file, and sent to DHRSC for analysis.

Six demographic variables are included in the survey of military health care beneficiaries. They are

1. SAMPLE GROUP
    (a) RETIRED - GREATER THAN 65 YEARS OLD
    (b) RETIRED - LESS THAN 65 YEARS OLD
    (c) ACTIVE DUTY (first sample only)

2. DEPENDENCY CODE
    (a) SPONSOR - non-active duty service member.
    (b) SPOUSE
    (c) CHILDREN
    (d) OTHER
3. GENDER
4. AGE
5. ZIP CODE
6. WHETHER OR NOT AN INDIVIDUAL HAS PRIVATE INSURANCE

For each type of insurance used (MTF, TriCare, or Private Insurer), respondents are asked the number of:

6. OVERNIGHT HOSPITAL STAYS
7. DOCTOR VISITS
8. PRESCRIPTIONS FILLED

In addition, participants are asked to subjectively evaluate their satisfaction with each of the three health care options (1 - Very Satisfied to 5 - Very Dissatisfied) and to rate their overall health (1 - Excellent to 5 - Poor). A sample Pre-Closing health care questionnaire is listed in Appendix B.

## E. ORGANIZATION OF THE DATA WITH UEDIT2

Survey data is stored in a text file format. The data was imported into UEDIT2. Because there was no white space column delimiters in the stored database, UEDIT2 imported the data in three sections - a row label, a character column, and a numeric column. The initial file is shown in figure 5.

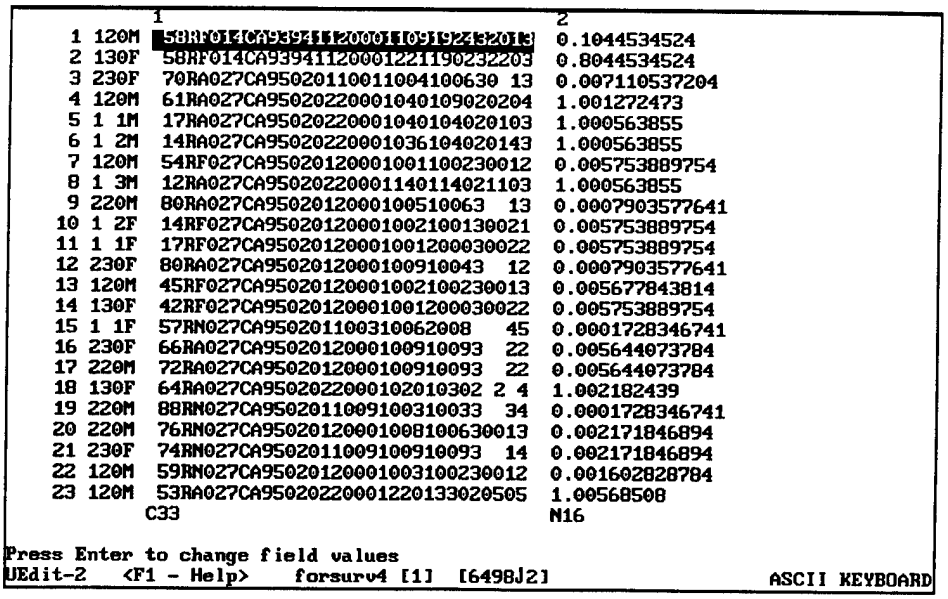



Figure 5. Data file as Imported into UEDIT2 before editing.

The first task was to restructure the data file so that it could be put in a form suitable for analysis. With the file description for the database text file format and using the ROW TOGGLE (Sh-F6), SPLIT COLUMN (Alt-F8), and EDIT COLUMN LABELS (Ctrl-F5) options in UEDIT2, the data was organized into separate columns. Further reduction was necessary to synchronize the two samples. Using the SHADOW command (Alt-F3) in UEDIT2, all active duty members and their dependents were removed from the first sample (no active duty responses where in the second sample). In addition, zip codes from the two samples had to filtered to match zip codes in the Monterey catchment area using the REDUCE (Ctrl F-8) function. The original database contained some responses from outside the sample being taken. Therefore, the REDUCE function allowed the author to cycle through each zip code and decide whether or not it was a zip code from the catchment area of interest. The data was sorted according to zip code and resaved in UEDIT2 format.

The organized file is shown in Figure 6. It should be noted that using the RECOLOR (Alt F-8) and SHADOW commands could have accomplished the same file format.

| | 1 SAMGRP | 2 DEPCODE | 3 GENDER | 4 AGE | 5 STATUS | 6 CATCH | 7 FORCE | 8 ZIPCODE | 9 INS | 10 HOSP | 1 H |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | M | 53 | R | 23 | A | 93901 | 2 | 2 | 0 |
| 2 | 1 | 1 | M | 61 | R | 23 | A | 93901 | 1 | 2 | 0 |
| 3 | 1 | 1 | M | 43 | R | 23 | F | 93901 | 2 | 2 | 0 |
| 4 | 1 | 1 | M | 49 | R | 23 | A | 93901 | 2 | 2 | 0 |
| 5 | 1 | 1 | M | 56 | R | 23 | A | 93901 | 2 | 2 | 0 |
| 6 | 1 | 1 | M | 64 | R | 23 | F | 93901 | 2 | 2 | 0 |
| 7 | 1 | 1 | M | 49 | R | 23 | A | 93901 | 1 | 2 | 0 |
| 8 | 1 | 1 | M | 64 | R | 23 | P | 93901 | 2 | 2 | 0 |
| 9 | 1 | 1 | M | 56 | R | 23 | A | 93901 | 1 | 1 | 0 |
| 10 | 1 | 1 | F | 36 | R | 23 | A | 93901 | 1 | 2 | 0 |
| 11 | 1 | 1 | M | 52 | R | 23 | A | 93901 | 1 | 2 | 0 |
| 12 | 1 | 1 | M | 64 | R | 23 | A | 93901 | 2 | 2 | 0 |
| 13 | 1 | 1 | M | 55 | R | 23 | N | 93905 | 2 | 2 | 0 |
| 14 | 1 | 1 | M | 64 | R | 23 | F | 93905 | 2 | 2 | 0 |
| 15 | 1 | 1 | M | 62 | R | 23 | A | 93905 | 2 | 2 | 0 |
| 16 | 1 | 1 | M | 55 | R | 23 | N | 93905 | 2 | 2 | 0 |
| 17 | 1 | 1 | M | 63 | R | 23 | A | 93905 | 2 | 2 | 0 |
| 18 | 1 | 1 | M | 63 | R | 23 | A | 93905 | 1 | 2 | 0 |
| 19 | 1 | 1 | M | 32 | R | 23 | A | 93905 | 2 | 2 | 0 |
| 20 | 1 | 1 | M | 54 | R | 23 | A | 93906 | 2 | 2 | 0 |
| 21 | 1 | 1 | M | 64 | R | 23 | A | 93906 | 1 | 2 | 0 |
| 22 | 1 | 1 | M | 57 | R | 23 | A | 93906 | 2 | 2 | 0 |
| | N6 | N7 | C6 | N3 | C6 | N5 | C5 | N7 | N3 | N4 | N |

Press Enter to change field values
UEdit-2 <F1 - Help> phase2a [1] [1321J27] ASCII KEYBOARD

Figure 6. Data file as Restructured in UEDIT2 After Editing and Sorting.

## F. RESPONSE

The first survey was conducted nationally in February, 1994 (before Silas B. Hayes closed) and the second survey was conducted for just the Fort Ord catchment in November, 1994 (after Silas B. Hayes closed). Participants were selected at random from the DEERS database [Ref. 3].

It should be noted that for purposes of this survey, Medicare was considered private insurance. The reason for this consideration in the survey is that military health care budgets do not cover retirees over age 65 on Medicare. Therefore, while a high percentage of respondents over age 65 indicate having private insurance, those people are budgeted through Medicare, not through military health care.

### 1. Pre-Closing Demographic Data

One thousand, seven hundred ninety-nine inquiries were mailed to eligible military health care beneficiaries in the catchment. Of these inquiries, 1,321 people replied (73.4 % response rate). Fifty-two percent of the respondents were age 65 or older. Seventy-nine percent of the respondents were male. The average age of the participants was 62.54 with the first quartile - age 56, the 2nd quartile - age 65, and the third quartile - age 72. Ninety-two percent of the respondents over age 65 and 46 % under age 65 have private health care insurance.

### 2. Post-Closing Demographic Data

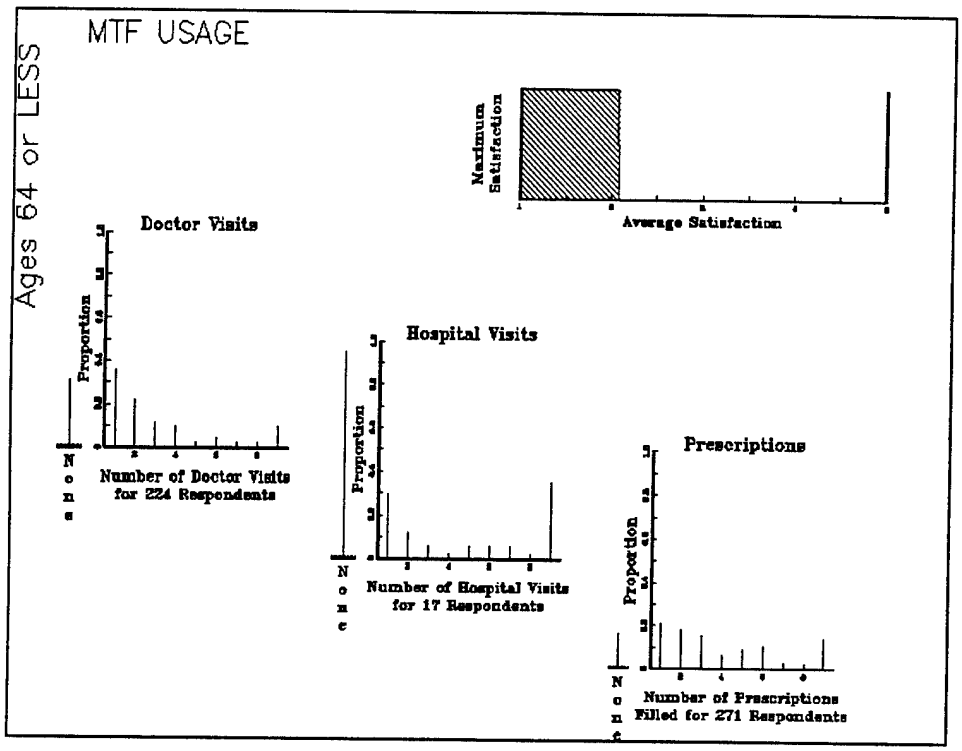
Thirteen thousand surveys were mailed to eligible military health care beneficiaries in the catchment. Of these inquiries, 6,293 people replied (48.4 % response rate). Fifty-one percent of the respondents were age 65 or older and 52 % of the respondents were male. The average age of the participants was 60.65 with the first quartile - age 54, the second quartile - age 64, and the 3rd quartile - age 71. Ninety-four percent of the respondents over age 65 and 49 % under age 65 have private health care insurance.

### 3. Display of Usage Patterns

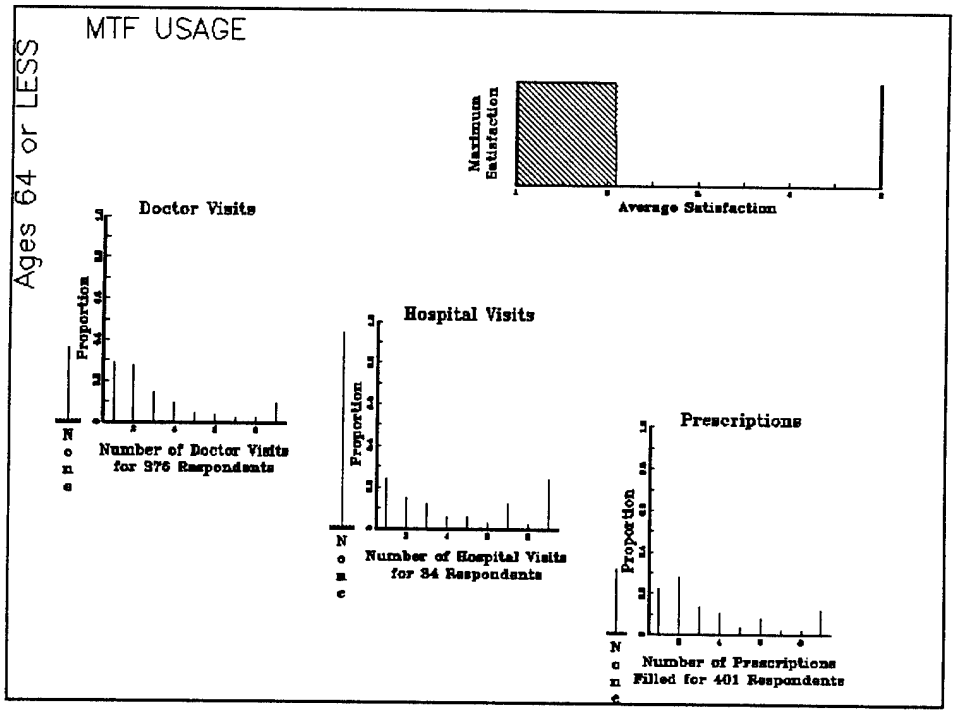
The desired graphic for displaying the military health care beneficiary data requires six variables - four independent variables and two dependent variables. The independent variables are Pre- and Post-Closing use, treatment source (MTF, TriCare, or Private), treatment type (doctor visit, hospital visit, and pharmacy use), and age. The two dependent variables are the number of occurrences in each treatment source for each treatment type (from 1 to 9) and the average satisfaction based on two of the independent variables (Pre/Post-Closing and age).

Display of the usage data is best illustrated by frequency counts. Using UEDIT2 to display the number of occurrences in each age group conditioned on having at least one response in each category (MTF, TriCare, Private), the variables were saved as APL2 matrices using the PUT (Sh-F3) command and then were imported into the APL2

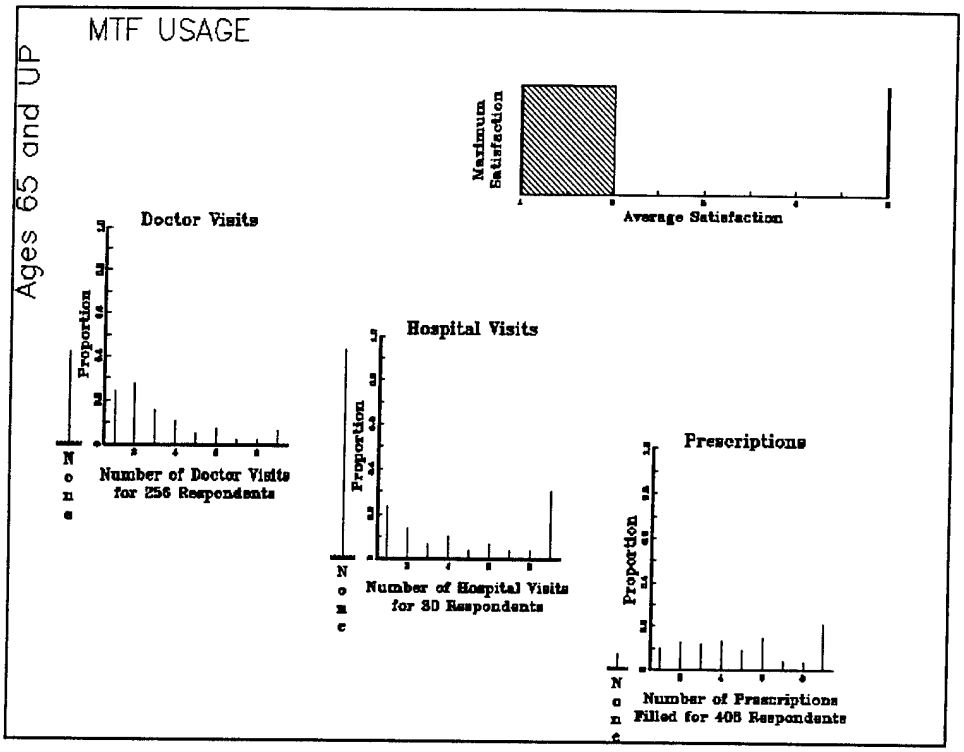
graphics package, AGSS. An APL2 program was written in AGSS to display each of the usage options for the two age groups. In addition, the average satisfaction of eligible users was also calculated to observe whether or not a significant shift in opinion had occurred between the two samples. The following graphs show how each age group within each sample used the medical facilities available to them:



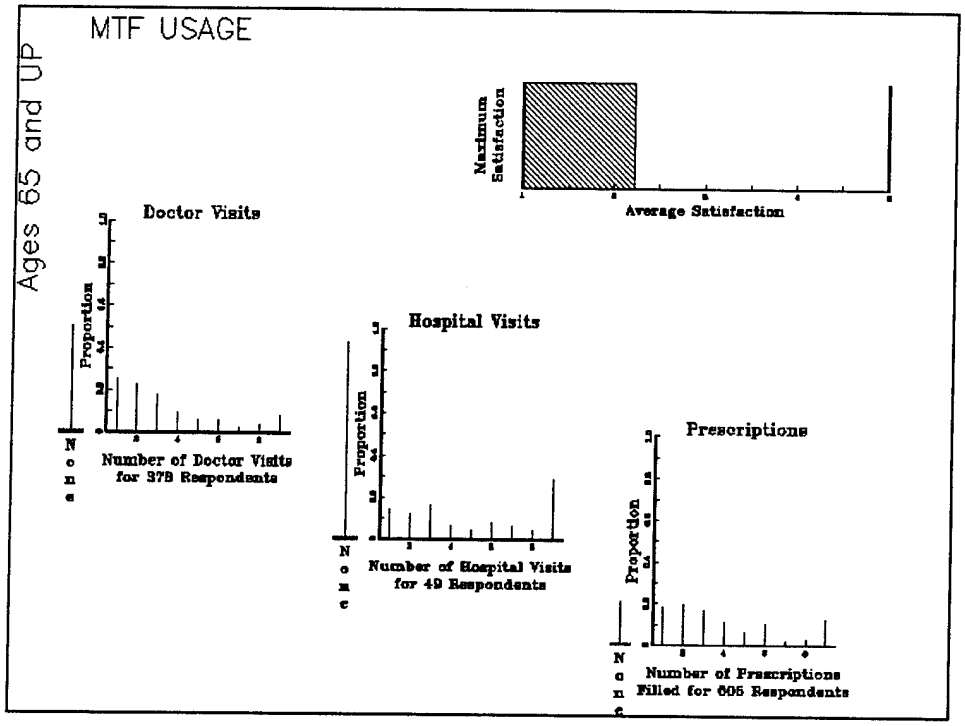

**Pre-Closing** MTF Usage by Beneficiaries Less Than 65 Years of Age.



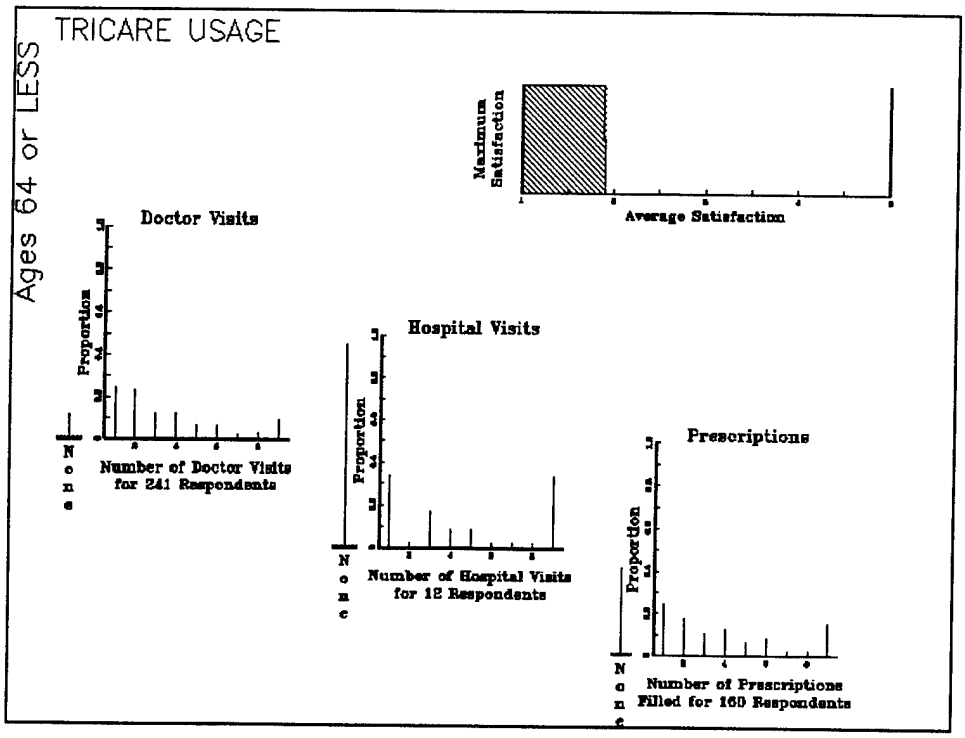

**Post-Closing** MTF Usage by Beneficiaries Less Than 65 Years of Age.



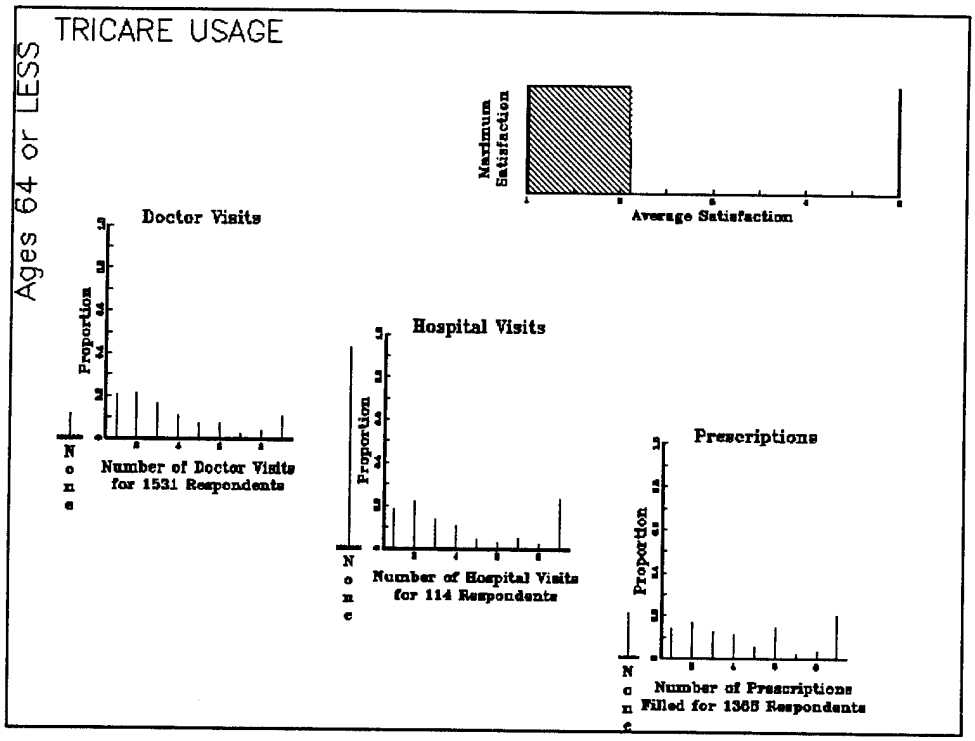

**Pre-Closing** MTF Usage by Beneficiaries Greater than 65 Years of Age.



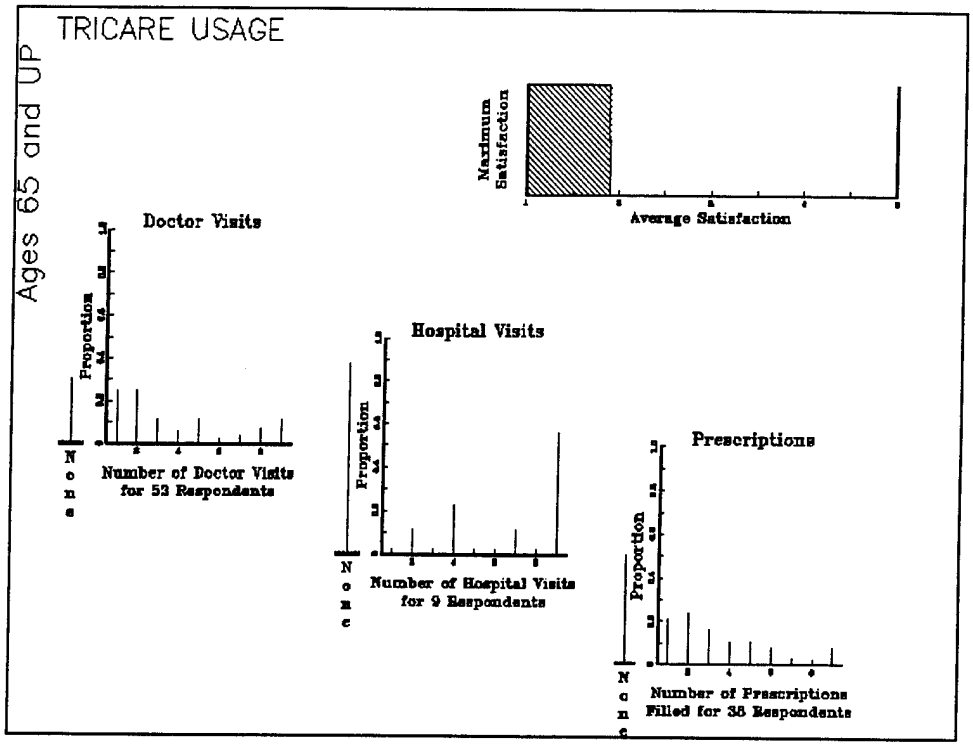

**Post-Closing** MTF Usage by Beneficiaries 65 Greater than 65 Years of Age.



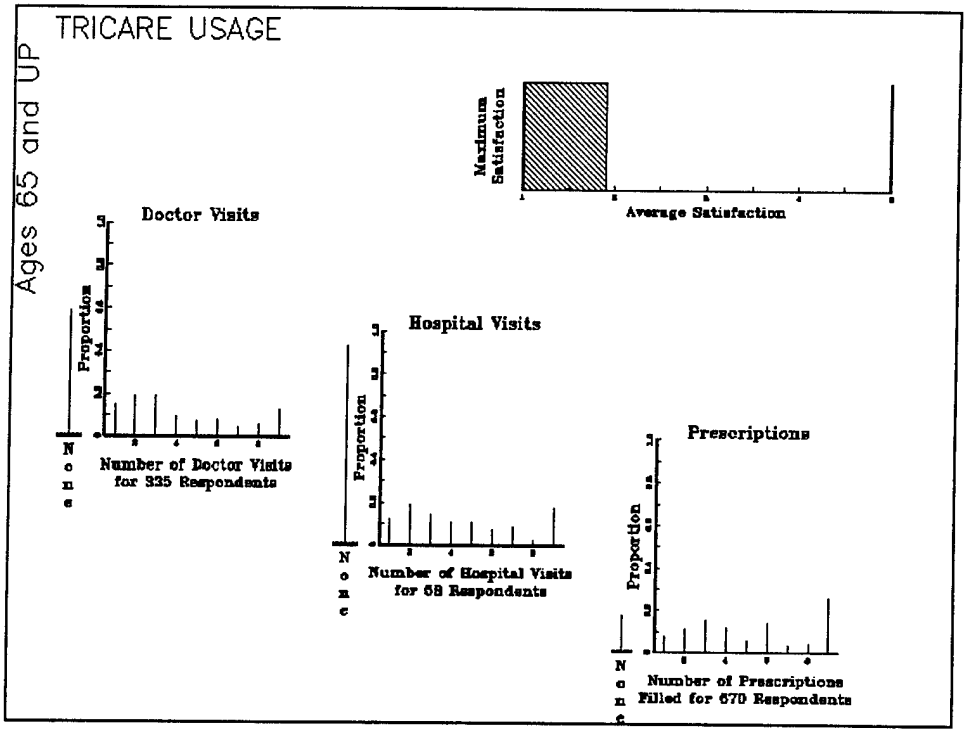

**Pre-Closing** TriCare Usage by Beneficiaries Less Than 65 Years of Age.



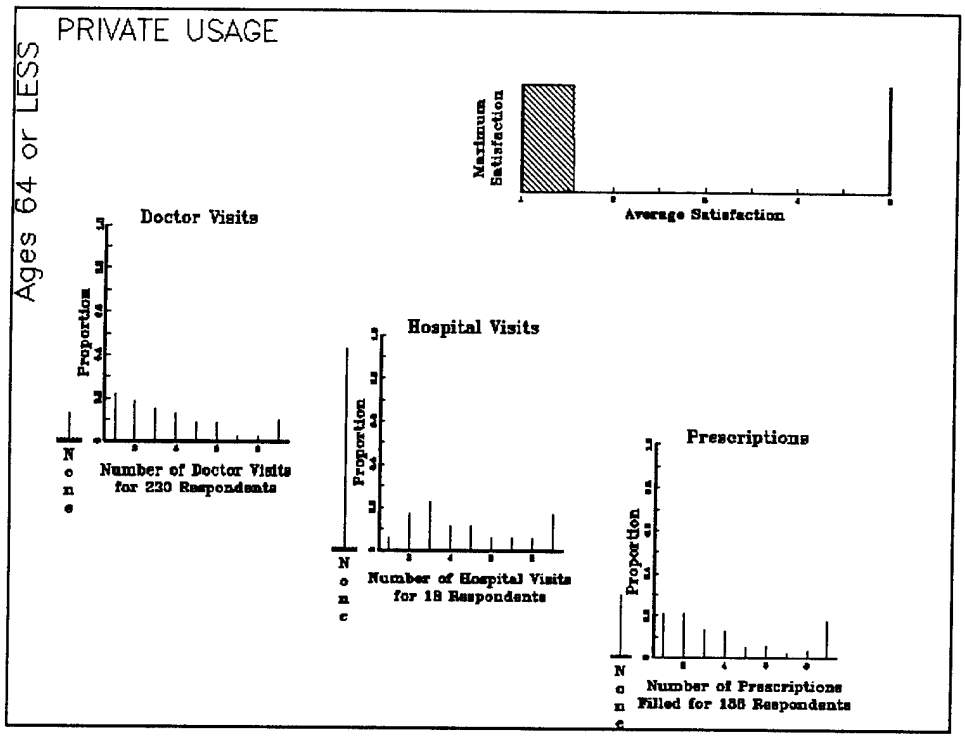

**Post-Closing** TriCare Usage by Beneficiaries Less Than 65 Years of Age.



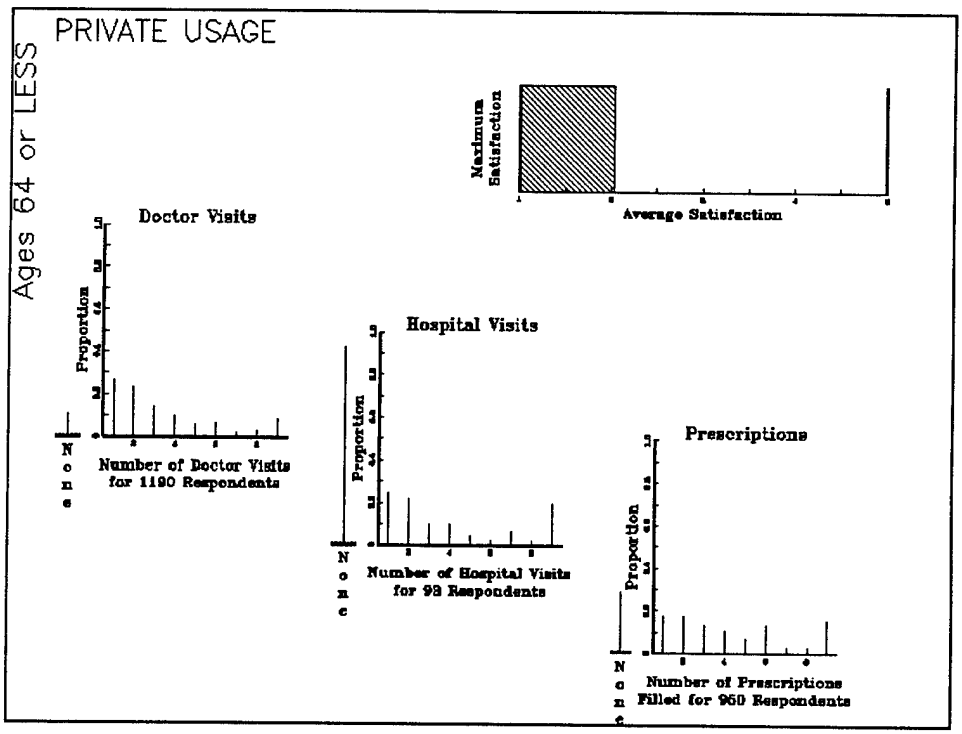

**Pre-Closing** TriCare Usage by Beneficiaries Greater Than 65 Years of Age.



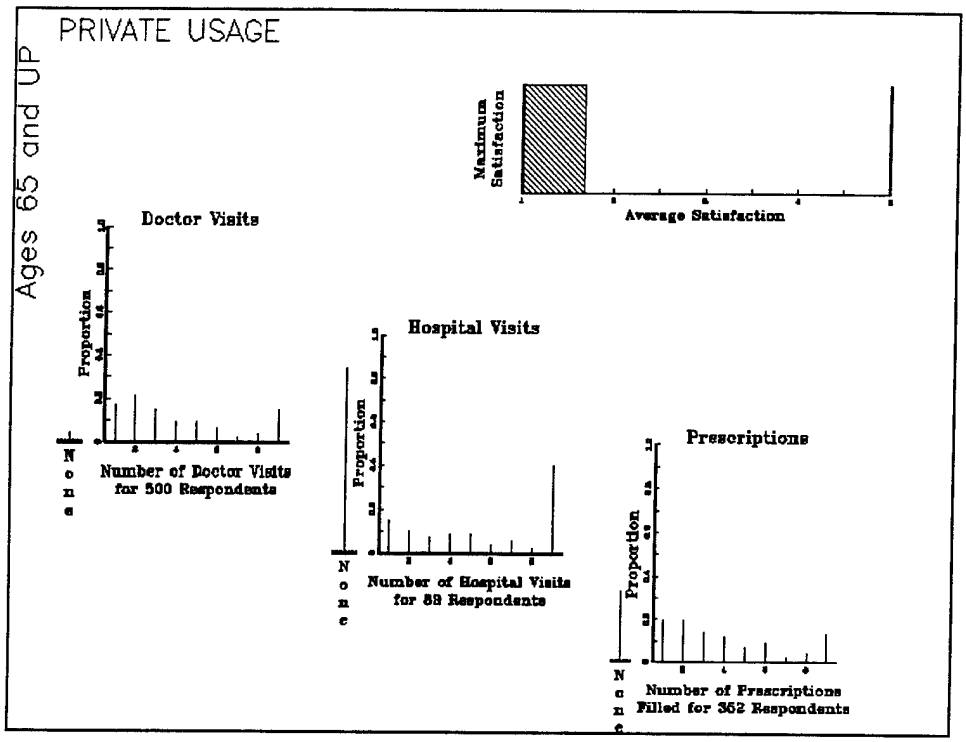

**Post-Closing** TriCare Usage by Beneficiaries Greater Than 65 Years of Age.



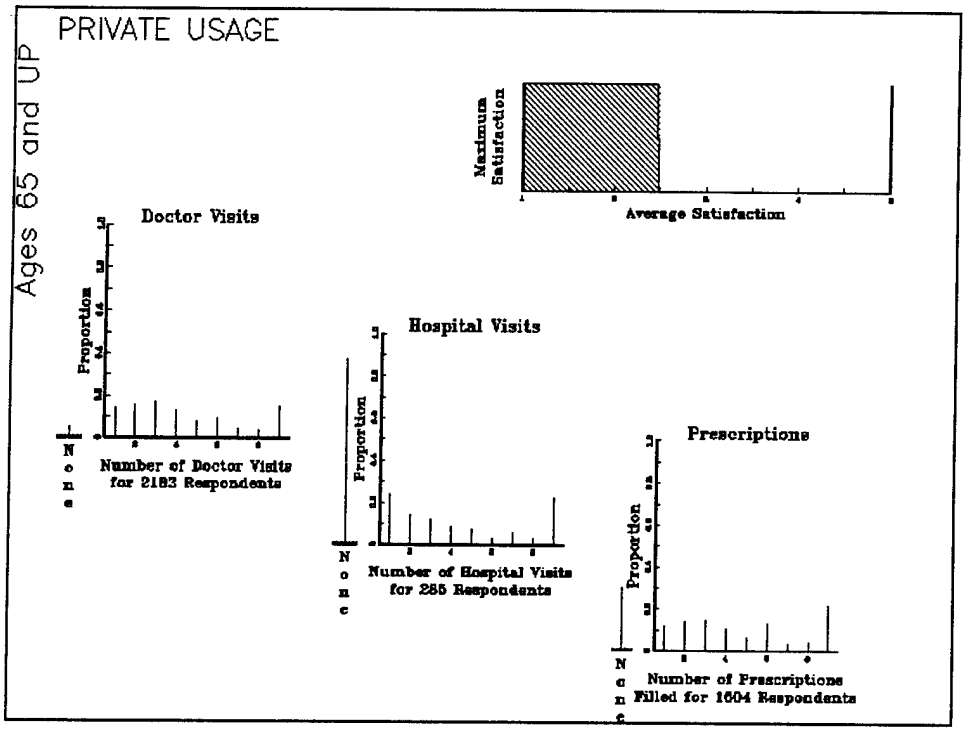

**Pre-Closing** Private Usage by Beneficiaries Less Than 65 Years of Age.




**Post-Closing** Private Usage by Beneficiaries Less Than 65 Years of Age.




**Pre-Closing** Private Usage by Beneficiaries Greater Than 65 Years of Age.




**Post-Closing** Private Usage by Beneficiaries Greater Than 65 Years of Age.

**4. Usage Shifts between the Two Samples**

***a. MTF Usage***

An overall decline in MTF use between samples occurred for each type of service surveyed in both age groups. This result is not surprising since using a MTF after Silas B. Hayes closed requires traveling to the nearest open MTF - most likely Oakland Naval Hospital.

***b. TriCare Usage***

An increase in TriCare use occurred between the two samples. Of particular interest is that an increase occurred in pharmacy use even for people over age 65. Ordinarily, people over 65 are not eligible for TriCare benefits due to coverage by Medicare. However, a special waiver was granted for prescription refills because of the hospital closure. People over age 65 are using this avenue for health care.

***c. Private Usage***

No real significant change in use occurred between the two samples. A slight decline in the average satisfaction level for both age groups is evident.

The percent of use and the change in percentage between the two samples is summarized in Table 7. It should be emphasized that use of the different alternatives is not mutually exclusive.

| | **Percentage Of Usage by Age Group** | | | | | | Change From Pre- to Post-Closing | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Pre-Closing** | | | **Post-Closing** | | | | | |
| Facility | < 65 | > 65 | Total | < 65 | > 65 | Total | < 65 | > 65 | Total |
| MTF | 51 % | 63 % | 57 % | 17 % | 25 % | 21 % | - 34 % | - 38 % | - 36 % |
| TriCare | 42 % | 11 % | 26 % | 52 % | 27 % | 40 % | +10 % | +16 % | +14 % |
| Private | 41 % | 75 % | 59 % | 40 % | 76 % | 57 % | - 1 % | + 1 % | - 2 % |

Table 7. Percent of Facility Usage for the Two Samples.

## 5. Conclusion

Four conclusions can be drawn from the graphical display of the usage patterns for the different facilities and age groups.

(i) A majority of the distributions for facility use are bimodal. Either a person went one or two times or more than nine times to each type of facility/service.

(ii) People did use Silas B. Hayes hospital quite extensively. A sharp decline in MTF use has occurred - presumably because of the distance to the nearest open MTF.

(iii) A significant increase in TriCare use has occurred, although not to the extent that there has been a drop in MTF use.

(iv) No significant change in use of private payment/insurance has occurred.

# APPENDIX A. STATISTICAL ANALYSIS FUNCTIONS AVAILABLE IN UEDIT2 AND EXCEL

UEDIT2's Statistical Menu

Frequency Count

Crosstabulations

Aggregation of Crostabulations

Box Plot

Plot X vs. Ys

Bootstrap Analysis for Contigency Tables

Multiple Regression - all combinations

Multiple Regression - specific combinations

Subpopulation Analysis

Excel's ToolPack Options

Anova: Single Factor

Anova: Two-Factor with Replication

Anova: Two-Factor without Replication

Correlation

Covariance

Descriptive Statistics

Exponential Smoothing

F-Test: Two-Sample for Variances

t-Test: Two-Sample Assuming Equal Variances

t-Test: Two-Sample Assuming Unequal Variances

t-Test: Paired Two-Sample for Means

z-test: Two-Sample for Means

Histogram

Moving Average

Random Number Generation

Rank and Percentile

Regression

Fourier Analysis

Sampling

# APPENDIX B. PRE-CLOSING HEALTH CARE QUESTIONNAIRE

## HEALTH CARE QUESTIONNAIRE

When completing the questions below, please keep the following definitions in mind.

**Military Medical Treatment Facility (MTF):** any military hospital or clinic. **It does not include VA facilities.**

**Civilian Hospital/Clinic:** any hospital or clinic open to the general civilian population.

**CHAMPUS:** includes CHAMPUS, CHAMPUS PRIME or EXTRA and any deductibles or copayments that you may personally pay.

**Private Payment:** any private medical insurance such as Blue Cross, Prudential or insurance that is paid by an employer or family member. Also includes the use of personal funds to pay for health care.

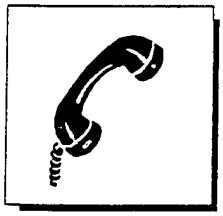
**Health care provider:** any doctor, nurse or other medical personnel.

### Phone Response

**Please respond by calling**

**1-800-883-4772**

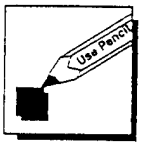
**Follow the instructions provided.**



### Mail Response

**If you choose to respond by mail, darken the square for each question and return your questionnaire in the postage paid envelope provided.**

- **Use a No. 2 lead pencil only.**
- **Do not use ink or ballpoint pen.**
- **Make heavy dark marks that fill the square.**
- **Erase cleanly any answer you wish to change.**
- **Make no stray marks.**




**IMPORTANT**

**Please respond to the following questions as they apply to the person to whom the questionnaire is addressed.**

1. Please enter the 7-digit access number that appears above the mailing address.

☐ ☐ ☐ ☐ ☐ ☐ ☐

2. Do you have any medical insurance/coverage other than CHAMPUS or CHAMPUS Supplemental Insurance? (This means medical insurance paid for by your employer or yourself or medical coverage such as MEDICARE/MEDICAID).

Yes [1] No [2]

3. During the past 6 months, did you stay in a hospital overnight?

Yes [1] No [2] → Go to Question 4

I spent the following number of nights in a:

| | None | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9+ |
|---|---|---|---|---|---|---|---|---|---|---|
| a. Military Hospital | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| b. Civilian hospital-CHAMPUS | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| c. Civilian hospital-private payment or other | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

446398

DO NOT WRITE IN THIS SHADED AREA

4. During the past 6 months, did you have an office visit with a doctor, nurse or other medical professional?

Yes [1] No [2] → Go to Question 5

I had the following number of office visits with the following health care provider(s) listed below:

| | None | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9+ |
|---|---|---|---|---|---|---|---|---|---|---|
| a. Military provider | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| b. Civilian provider-CHAMPUS | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| c. Civilian provider-private payment or other | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

5. During the past 6 months, did you fill or refill any prescriptions?

Yes [1] No [2] → Go to Question 6

I filled or refilled the following number of prescriptions at the pharmacy(s) listed below:

| | None | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9+ |
|---|---|---|---|---|---|---|---|---|---|---|
| a. Military pharmacy | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| b. Civilian pharmacy-CHAMPUS | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| c. Civilian pharmacy-private payment or other | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

6. Overall, how satisfied are you with the health care services you have received during the last 6 months from a:

| | Does Not Apply | Very Satis-fied | Satis-fied | Neither | Dissat-isfied | Very Dissat-isfied |
|---|---|---|---|---|---|---|
| a. Military health care facility/provider | 0 | 1 | 2 | 3 | 4 | 5 |
| b. Civilian facility/provider paid by CHAMPUS | 0 | 1 | 2 | 3 | 4 | 5 |
| c. Civilian facility/provider paid by private payment | 0 | 1 | 2 | 3 | 4 | 5 |

7. Overall, how would you describe your health?

Excellent [1] Very Good [2] Good [3] Fair [4] Poor [5]

8. Please enter your zip code (if you are completing this survey for a family member, enter their zip code).

[ ] [ ] [ ] [ ] [ ]

NOTE: If responding by mail, please fill in this grid.

Write the zip code in the boxes. →

Then fill in the matching circles below each box. →

| | | | | |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 | 4 |
| 5 | 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 | 6 |
| 7 | 7 | 7 | 7 | 7 |
| 8 | 8 | 8 | 8 | 8 |
| 9 | 9 | 9 | 9 | 9 |

THIS COMPLETES THE QUESTIONNAIRE - PLEASE RESPOND BY FEBRUARY 18.

R8168-PFI-54321

# APPENDIX C. APL2 CODE FOR SURVEY ANALYSIS

```
        APL2 CODE FOR ANALYSIS OF HEALTH CARE PREFERENCES

PAGE 1          1995-03-10   10.31.47


*****GENERALPR      1995-03-10  10.16.05     3 ⎕AT = 0 0 0 0
     [0]    GENERALPR;TEMPDX;TEMPDY;TEMPHX;TEMPHY;TEMPPX;TEMPPY;TEMPSX;TEMPSY
     [1]    ⍝⍝⍝ DO NOT MOVE OR ERASE; AGSS FUNCTION HEADER
     [2]    ⍝⍝⍝      AGSS WILL NOT ADD A LINE TO THIS FUNCTION WITHOUT THIS HEADER
     [3]     ERASE
     [4]
     [5]    ⍝
     [6]    ⍝
     [7]     TOPNAME3←'PRIVATE USAGE'
     [8]     TOPNAME2←'TRICARE USAGE'
     [9]     TOPNAME1←'MTF USAGE'
     [10]    SIDENAME1←'Ages 65 and UP'
     [11]    SIDENAME2←'Ages 64 or LESS'
     [12]    TOPNAME←∈((J=1)(J=2)(J=3))/TOPNAME1 TOPNAME2 TOPNAME3
     [13]    SIDENAME←∈((I=1)(I=2))/SIDENAME2 SIDENAME1
     [14]    XDL1←0.1
     [15]    XDR1←0.3
     [16]    XDL10←0.05
     [17]    XDR10←0.075
     [18]    YDL1←0.4
     [19]    YDU1←0.7
     [20]   ⍝
     [21]    XHL1←0.4
     [22]    XHR1←0.6
     [23]    XHL10←0.35
     [24]    XHR10←0.375
     [25]    YHL1←0.25
     [26]    YHU1←0.55
     [27]   ⍝
     [28]    XPL1←0.7
     [29]    XPR1←0.9
     [30]    XPL10←0.655
     [31]    XPR10←0.675
     [32]    YPL1←0.1
     [33]    YPU1←0.4
     [34]   ⍝
     [35]    XSL1←0.55
     [36]    XSR1←0.95
     [37]    YSL1←0.75
     [38]    YSU1←0.9
     [39]   ⍝
     [40]    TEMPPX←(⍎(⍎VARIJP),'[;1]')
     [41]    TEMPPY←(⍎(⍎VARIJP),'[;2]')
     [42]    TEMPDX←(⍎(⍎VARIJD),'[;1]')
     [43]    TEMPDY←(⍎(⍎VARIJD),'[;2]')
     [44]    TEMPHX←(⍎(⍎VARIJH),'[;1]')
     [45]    TEMPHY←(⍎(⍎VARIJH),'[;2]')
     [46]    TEMPSX←(⍎(⍎VARIJS),'[;1]')
     [47]    TEMPSY←(⍎(⍎VARIJS),'[;2]')
     [48]   ⍝
```

```
*****GENERALPR      1995-03-10  10.16.05     3 ⎕AT = 0 0 0 0   CONT'D
     [49]   DAYHOSP←'Days in Hospital'
     [50]   DOCVIS←'No Doctor Visits'
     [51]   PERC←'Percentage'
     [52]   PRFIL←'Prescriptions Filled'
     [53]   rUN AXISCOLOR
     [54]   rUN GENERALD
     [55]   rUN GENERALD0
     [56]   rUN D0AXIS
     [57]   rUN GENERALH0
     [58]   rUN GENERALH
     [59]   rUN H0AXIS
     [60]   rUN GENERALP
     [61]   rUN GENERALP0
     [62]   rUN P0AXIS
     [63]   rUN SATIS
     [64]   rUN TAXIS
     [65]   rUN TOPTITLE
     [66]   rUN SIDETITLE
     [67]   VIEW
     [68]
     [69]  ⍝THE FOLLOWING VARIABLES USED IN THE PROGRAM ARE DEFINED AS FOLLOWS:
     [70]  ⍝VARIJP
     [71]  ⍝ 'S165',(((I=1))/'MP'),(((J=1),(J=2),(J=3))/'MTP'),'P'
     [72]  ⍝VARIJH
     [73]  ⍝ 'S165',(((I=1))/'MP'),(((J=1),(J=2),(J=3))/'MTP'),'H'
     [74]  ⍝VARIJD
     [75]  ⍝ 'S165',(((I=1))/'MP'),(((J=1),(J=2),(J=3))/'MTP'),'D'
     [76]  ⍝VARIJS
     [77]  ⍝ 'S165',(((I=1))/'MP'),(((J=1),(J=2),(J=3))/'MTP'),'S'
     [78]  ⍝
     [79]  ⍝      The variables S(1/2)65(M/P)(M/T/P)(P/H/D/S) refer to the
     [80]  ⍝raw numbers for the frequency counts taken in UEDIT2 from the data
     [81]  ⍝files (phase2a.UED and phase4a.UED).  For example, S165MMP refers
     [82]  ⍝to the count for pre-closing data (S1), people ages or less (65M),
     [83]  ⍝military treatment facility (M), number of prescriptions filled (P).
```

## LIST OF REFERENCES


1. Steinfeld, CAPT Uwe H., UEDIT 1.09, An APL2 Input/Output Editor and Exploratory Data Analysis Tool, Master's Thesis, Naval Postgraduate School, Monterey, California, September 1992.

2. Microsoft Excel User's Guide, Microsoft Corporation, 1994.

3. MHSS Beneficiary Survey - National Implementation, Ernst & Young, July, 1994.

## INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center ..... 2
   Cameron Station
   Alexandria, Virginia 22304-6145

2. Library, Code 52 ..... 2
   Naval Postgraduate School
   Monterey, California 93943-5101

3. Professor Peter A.W. Lewis, Code OR/LW ..... 1
   Department of Operations Research
   Naval Postgraduate School
   Monterey, California 93940-5000

4. Professor William Kemple, Code OR/KE ..... 1
   Department of Operations Research
   Naval Postgraduate School
   Monterey, California 93940-5000

5. Captain James Scaramozzino, MSC, USN ..... 1
   Defense Health Resources Study Center
   Naval Postgraduate School, Code 65
   Monterey, California 93940

6. Lieutenant John H. Barnet Jr. ..... 1
   # 2 Robert S. Drive
   Menlo Park, California 94025